

**ANALISIS SENTIMEN PEMUTIHAN PAJAK KENDARAAN  
BERMOTOR DI APLIKASI X DENGAN MODEL HIBRIDA  
INDOBERT-CNN LSTM**

**SKRIPSI**

**Oleh:**

**LOUIS BRANDON**

**NIM. 22113763**



**PROGRAM STUDI S-1 TEKNIK INFORMATIKA  
FAKULTAS INFORMATIKA  
UNIVERSITAS MIKROSKIL  
MEDAN  
2025**

# **SENTIMENT ANALYSIS IN X ON VEHICLE TAX AMNESTY WITH INDOBERT-CNN LSTM HYBRID MODEL**

**UNDERGRADUATE THESIS**

**By:**

**LOUIS BRANDON**

**ID NUMBER. 22113763**



**UNDERGRADUATE PROGRAM OF COMPUTER SCIENCE  
FACULTY OF INFORMATICS  
UNIVERSITAS MIKROSKIL  
MEDAN  
2025**

LEMBARAN PENGESAHAN

**Analisis Sentimen Pemutihan Pajak Kendaraan Bermotor di Aplikasi X  
dengan Model Hibrida IndoBERT-CNN LSTM  
SKRIPSI**

**SKRIPSI**

Diajukan untuk Melengkapi Persyaratan Guna  
Mendapatkan Gelar Sarjana  
Program Studi S-1 Teknik Informatika

Oleh:

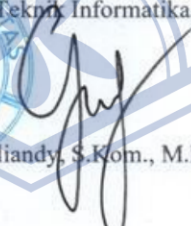
**LOUIS BRANDON**  
**NIM. 22113763**

Disetujui Oleh:

|                                                                                     |                                                                                      |
|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| Dosen Pembimbing I,                                                                 | Dosen Pembimbing II,                                                                 |
|  |  |
| Sunaryo Winardi, S.Kom., M.T                                                        | R. A. Fattah Adriansyah, S.Kom., M.Kom.                                              |

Medan, 11 Agustus 2026  
Diketahui dan Disahkan Oleh:

Ketua Program Studi  
S-1 Teknik Informatika,

  
Charles Juliandy, S.Kom., M.Kom.

## HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-1 Teknik Informatika Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Louis Brandon

NIM : 221113763

Saya telah melaksanakan penelitian dan penulisan Skripsi dengan judul dan tempat penelitian sebagai berikut:

Judul Skripsi : Analisis Sentimen Pemutihan Pajak Kendaraan Bermotor di Aplikasi X dengan Model Hibrida IndoBERT-CNN LSTM

Sehubungan dengan Skripsi tersebut, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Skripsi tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Skripsi saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Skripsi saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Skripsi saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 25 Juli 2026

Saya yang membuat pernyataan,



Louis Brandon

# **Analisis Sentimen Pemutihan Pajak Kendaraan Bermotor di Aplikasi X dengan Model Hibrida IndoBERT-CNN LSTM**

## **Abstrak**

*Kebijakan pemutihan pajak kendaraan bermotor banyak diperbincangkan masyarakat di media sosial X, mencerminkan beragam opini yang perlu dianalisis untuk memahami persepsi publik. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap kebijakan tersebut menggunakan model hibrida IndoBERT-CNN-LSTM. IndoBERT digunakan untuk menghasilkan representasi kata kontekstual pada teks berbahasa Indonesia, dilanjutkan dengan CNN untuk mengekstraksi fitur lokal signifikan dan LSTM untuk menangkap dependensi sekuensial serta konteks jangka panjang. Data diperoleh melalui crawling tweet dengan kata kunci terkait pemutihan pajak, kemudian melalui tahap prapemrosesan (cleaning, case folding, tokenizing, stopword removal, stemming) sebelum dilabeli ke dalam tiga kelas sentimen: positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa model hibrida IndoBERT-CNN-LSTM mampu meningkatkan akurasi klasifikasi sentimen dibandingkan model tunggal, dengan memanfaatkan keunggulan IndoBERT dalam pemahaman konteks bahasa serta kemampuan CNN dan LSTM dalam ekstraksi fitur dan pemodelan sekuensial. Hasil ini diharapkan memberikan gambaran persepsi publik terhadap kebijakan pemutihan pajak kendaraan bermotor sekaligus menjadi bahan evaluasi pemerintah.*

**Kata kunci:** analisis sentimen, pemutihan pajak kendaraan bermotor, IndoBERT, CNN-LSTM, media sosial

## **Abstract**

*The vehicle tax amnesty policy is widely discussed on social media platform X, reflecting diverse public opinions that need analysis to understand public perception. This study analyzes public sentiment toward the policy using a hybrid IndoBERT-CNN-LSTM model. IndoBERT generates contextual word representations for Indonesian text, followed by CNN to extract significant local features and LSTM to capture sequential dependencies and long-term context. Data were obtained through crawling tweets using keywords related to the tax amnesty policy, then preprocessed (cleaning, case folding, tokenizing, stopword removal, stemming) before being labeled into three sentiment classes: positive, negative, and neutral. Results show that the hybrid IndoBERT-CNN-LSTM model improves sentiment classification accuracy compared to single models, leveraging IndoBERT's contextual language understanding alongside CNN and LSTM's feature extraction and sequential modeling capabilities. These findings are expected to provide insight into public perception of the vehicle tax amnesty policy and serve as evaluation material for the government.*

**Keywords:** sentiment analysis, vehicle tax amnesty, IndoBERT, CNN-LSTM, social media

## KATA PENGANTAR

Ucapan syukur saya panjatkan terhadap Tuhan Yang Maha Esa atas segala rahmat, karunia, dan kemudahan yang diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada program studi yang saya tempuh.

Penulisan skripsi ini dilatarbelakangi oleh ketertarikan saya terhadap pelaksanaan program pemutihan pajak kendaraan yang diterapkan pemerintah daerah, serta menilai kontribusi baik secara teoritis maupun praktis melalui pendapat yang diutarakan oleh masyarakat dari media sosial. Saya berharap hasil penelitian ini dapat menambah wawasan, memberikan solusi atas permasalahan yang dibahas, serta bermanfaat bagi pembaca, instansi terkait, maupun penelitian selanjutnya.

Dalam proses penyusunan skripsi ini, saya banyak mendapatkan bimbingan, dukungan, serta bantuan dari berbagai pihak. Oleh karena itu, saya ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Sunaryo Winardi, S.Kom., M.T, selaku Dosen Pembimbing I dan Dekan Fakultas Informatika Universitas Mikroskil Medan.
2. Bapak R. A. Fattah Adriansyah, S.Kom., M.Kom., selaku Dosen Pembimbing II.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D, selaku Rektor Universitas Mikroskil Medan.
4. Bapak Carles Juliandy, S.Kom., M.Kom., selaku Ketua Program Studi S-1 Fakultas Informatika Universitas Mikroskil Medan.
5. Orang tua dan keluarga

Saya menyadari bahwa skripsi ini masih memiliki beberapa kekurangan, baik dari segi penyusunan, analisis data, maupun keterbatasan referensi yang digunakan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang.

Medan, 25 Juli 2026

Penulis,

Louis Brandon

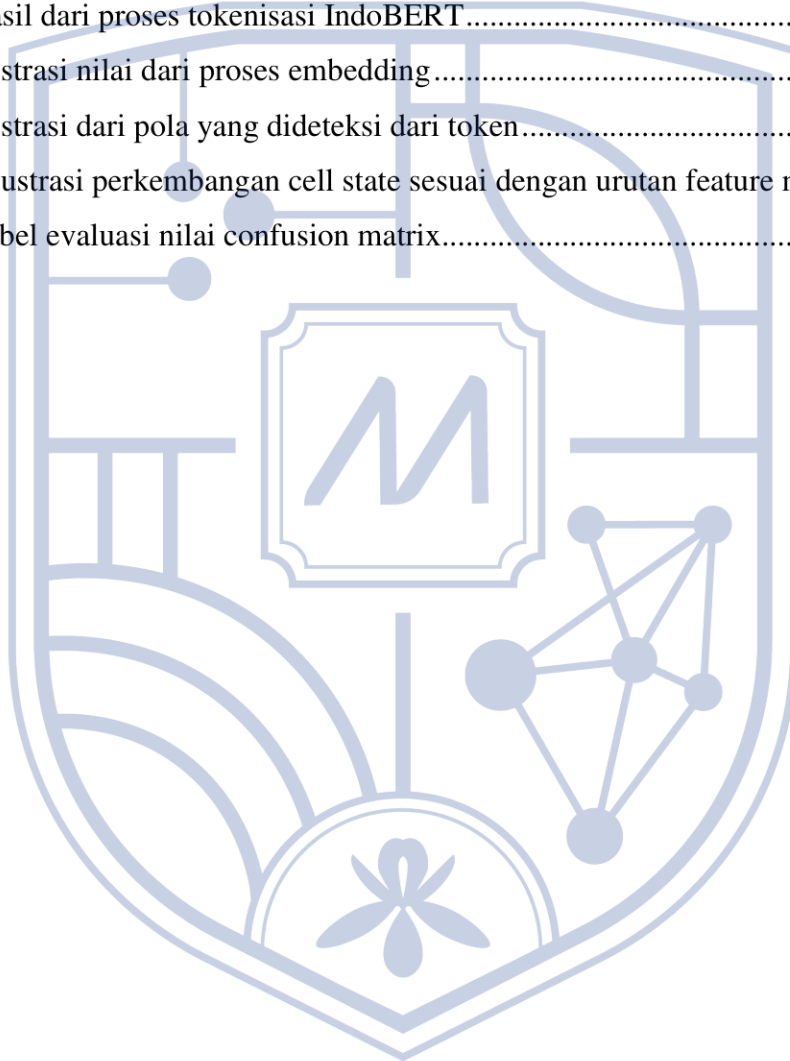
## DAFTAR ISI

|                                                                                                                        |            |
|------------------------------------------------------------------------------------------------------------------------|------------|
| <b>ANALISIS SENTIMEN PEMUTIHAN PAJAK KENDARAAN BERMOTOR DI APLIKASI X DENGAN MODEL HIBRIDA INDOBERT-CNN LSTM .....</b> | <b>I</b>   |
| <b>ABSTRAK.....</b>                                                                                                    | <b>I</b>   |
| <b>KATA PENGANTAR .....</b>                                                                                            | <b>II</b>  |
| <b>DAFTAR ISI .....</b>                                                                                                | <b>III</b> |
| <b>DAFTAR TABEL .....</b>                                                                                              | <b>V</b>   |
| <b>DAFTAR GAMBAR .....</b>                                                                                             | <b>VI</b>  |
| <b>BAB I PENDAHULUAN .....</b>                                                                                         | <b>1</b>   |
| 1.1. Latar Belakang .....                                                                                              | 1          |
| 1.2. Rumusan Masalah .....                                                                                             | 2          |
| 1.3. Tujuan .....                                                                                                      | 3          |
| 1.4. Manfaat .....                                                                                                     | 3          |
| 1.5. Ruang Lingkup.....                                                                                                | 3          |
| <b>BAB II KAJIAN LITERATUR.....</b>                                                                                    | <b>4</b>   |
| 2.1. Pemutihan Pajak.....                                                                                              | 4          |
| 2.2. Analisis Sentimen .....                                                                                           | 4          |
| 2.3. <i>Natural Language Processing (NLP)</i> .....                                                                    | 6          |
| 2.4. Machine Learning .....                                                                                            | 7          |
| 2.5. <i>Transformer</i> dan BERT .....                                                                                 | 8          |
| 2.6. IndoBERT .....                                                                                                    | 9          |
| 2.7. <i>Convolution Neural Network</i> .....                                                                           | 11         |
| 2.8. Operasi <i>Pooling</i> .....                                                                                      | 12         |
| 2.9. <i>Long Short Term Memory (LSTM)</i> .....                                                                        | 13         |
| 2.9.1. Forget Gate.....                                                                                                | 13         |
| 2.9.2. Input Gate.....                                                                                                 | 13         |
| 2.9.3. Cell State .....                                                                                                | 14         |
| 2.9.4. Output Gate & Hidden Gate.....                                                                                  | 14         |
| 2.10. Model Hibrida.....                                                                                               | 14         |
| 2.11. <i>Confusion Matrix</i> .....                                                                                    | 15         |
| <b>BAB III TAHAPAN PELAKSANAAN .....</b>                                                                               | <b>17</b>  |
| 3.1. Kerangka Tahapan Pelaksanaan.....                                                                                 | 17         |
| 3.2. Analisis Proses .....                                                                                             | 19         |

|                                                                                                                     |           |
|---------------------------------------------------------------------------------------------------------------------|-----------|
| 1.2.1. Input Dataset .....                                                                                          | 19        |
| 1.2.2. Pre-Processing Data .....                                                                                    | 20        |
| 1.2.3. Pelabelan Data.....                                                                                          | 22        |
| 3.3. Pembangunan Model IndoBERT .....                                                                               | 23        |
| 3.3.1. Tokenisasi IndoBERT .....                                                                                    | 24        |
| 3.3.2. Embedding IndoBERT .....                                                                                     | 25        |
| 3.4. Ekstraksi Fitur dengan CNN .....                                                                               | 29        |
| 3.4.1. Operasi Konvolusi 1D.....                                                                                    | 29        |
| 3.4.2. Ilustrasi Deteksi Pola Sentimen oleh Filter.....                                                             | 29        |
| 3.4.3. MaxPooling Temporal .....                                                                                    | 30        |
| 3.4.4. Dropout Regulisasi Pertama.....                                                                              | 30        |
| 3.4.5. Pemodelan Sekuensial oleh LSTM .....                                                                         | 30        |
| 3.4.6. Penjabaran Komputasi LSTM pada Kalimat S1 .....                                                              | 30        |
| 3.4.7. Evolusi Cell State pada S1 .....                                                                             | 31        |
| <b>BAB IV HASIL &amp; PEMBAHASAN .....</b>                                                                          | <b>32</b> |
| 4.1. Hasil .....                                                                                                    | 32        |
| 4.2. Tuning Model.....                                                                                              | 32        |
| 4.3.1. Evaluasi Hasil Model.....                                                                                    | 32        |
| 4.3.2. Analisis Sentimen Masyarakat terhadap Program Pemutihan Pajak<br>Kendaraan Bermotor di Aplikasi X .....      | 35        |
| 4.4. Evaluasi Model Hibrida Analisis Sentimen IndoBERT CNN-LSTM untuk<br>Klasifikasi Teks Berbahasa Indonesia ..... | 35        |
| <b>BAB V PENUTUP .....</b>                                                                                          | <b>36</b> |
| 5.1 Kesimpulan .....                                                                                                | 36        |
| 5.2 Saran .....                                                                                                     | 37        |
| <b>DAFTAR PUSTAKA .....</b>                                                                                         | <b>38</b> |

## DAFTAR TABEL

|                                                                                    |    |
|------------------------------------------------------------------------------------|----|
| Tabel 3.1 Hasil dari proses text cleaning.....                                     | 20 |
| Tabel 3.2 Hasil dari proses case folding.....                                      | 21 |
| Tabel 3.3 Hasil dari proses normalisasi.....                                       | 22 |
| Tabel 3.4 Contoh kata yang ditemukan dan nilainya dari kamus Inset.....            | 23 |
| Tabel 3.5 Hasil dari labelling data.....                                           | 23 |
| Tabel 3.6 Token khusus dari IndoBERT dan fungsinya .....                           | 24 |
| Tabel 3.7 Hasil dari proses tokenisasi IndoBERT.....                               | 25 |
| Tabel 3.8 Ilustrasi nilai dari proses embedding.....                               | 26 |
| Tabel 3.9 Ilustrasi dari pola yang dideteksi dari token.....                       | 30 |
| Tabel 3.10 Ilustrasi perkembangan cell state sesuai dengan urutan feature map..... | 31 |
| Tabel 4.1 Tabel evaluasi nilai confusion matrix.....                               | 34 |



## DAFTAR GAMBAR

|                                                                                  |    |
|----------------------------------------------------------------------------------|----|
| Gambar 3.1 Flowchart analisis proses pengerjaan penelitian.....                  | 19 |
| Gambar 4.1 <i>Confusion matrix</i> model setelah dilakukan <i>training</i> ..... | 34 |

