IMPLEMENTASI MODEL INDOBERT UNTUK MENANGANI KETIDAKSEIMBANGAN DATA DALAM ANALISIS SENTIMEN ULASAN APLIKASI KAI ACCESS

DYAH LISTYOWATI

PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul "Implementasi Model IndoBERT Untuk Menangani Ketidakseimbangan Data Dalam Analisis Sentimen Ulasan Aplikasi KAI Access" 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, April 2026

Dyah Listyowati
G14190061

ABSTRAK

DYAH LISTYOWATI. Implementasi Model Indobert Untuk Menangani Ketidakseimbangan Data Dalam Analisis Sentimen Ulasan Aplikasi KAI Access. Dibimbing oleh ANWAR FITRIANTO dan AAM ALAMUDI.

Analisis sentimen adalah proses pengolahan data tekstual untuk meneliti pendapat atau opini mengenai entitas tertentu. Analisis sentimen dapat diterapkan dalam berbagai hal, salah satunya pada aplikasi KAI Access. Penelitian ini menggunakan 4359 data ulasan aplikasi KAI Access dari bulan Juni 2025-Agustus 2025 yang diklasifikasikan ke kelas positif, negatif dan netral. Penelitian akan mengimplementasikan IndoBERT, salah satu model *Bidirectional Encoder Representation from Transformers* (BERT) monolingual dengan penerapan *Synthetic Minority Over-sampling Technique* dan *Random Oversampling* sebagai metode penanganan yang sering digunakan untuk menangani data tak seimbang. Tujuan dari penelitian ini adalah mengimplementasikan metode untuk penanganan data tak seimbang dan membandingkan model untuk menangani ketidakseimbangan sentimen ulasan pengguna pada aplikasi KAI Access. Data akan dibagi menjadi tiga yaitu 80% data latih, 10% data validasi dan 10% data uji. Penelitian ini menggunakan tiga skenario yaitu IndoBERT, IndoBERT dengan SMOTE, dan IndoBERT dengan ROS. Dari ketiga skenario tersebut, akurasi yang dihasilkan oleh model IndoBERT tanpa penanganan memiliki hasil yang terbaik namun evaluasi mendalam menunjukan bahwa model dengan SMOTE memiliki performa yang baik untuk kelas minor pada kasus data tak seimbang.

Kata Kunci: Analisis Sentimen, IndoBERT, KAI Access, Random Oversampling, SMOTE

ABSTRACT

DYAH LISTYOWATI. Implementation of the IndoBERT Model for Handling Data Imbalance in Sentiment Analysis of KAI Access Application Reviews. Supervised by ANWAR FITRIANTO dan AAM ALAMUDI.

Sentiment analysis is the process of processing textual data aimed at examining opinions or sentiments regarding a specific entity. It can be applied in various ways, such as in the KAI Access application. The dataset consists of 4.359 reviews collected from June 2025 to August 2025, categorized into positive, negative, and neutral classes. The research implements IndoBERT, a monolingual version of the Bidirectional Encoder Representation from Transformers (BERT) model, along with the application of *Synthetic Minority Over-sampling Technique* and *Random Oversampling* as commonly used methods to handle imbalanced data. This research identifies the sentiment of application user reviews and implements methods to handle imbalanced data. The data was partitioned into training (80%), validation (10%), and testing (10%) sets. The research evaluated three scenarios: a baseline IndoBERT model, IndoBERT with SMOTE, and IndoBERT with Random Oversampling. Among the three scenarios, the accuracy produced by the IndoBERT model without any imbalance handling yielded the best results. However, an in-depth evaluation indicates that the model utilizing SMOTE performs better for the minority class in cases of imbalanced data.

Keywords: IndoBERT, KAI Access, Random Oversampling, Sentiment Analysis, SMOTE

IMPLEMENTASI MODEL INDOBERT UNTUK MENANGANI KETIDAKSEIMBANGAN DATA DALAM ANALISIS SENTIMEN ULASAN APLIKASI KAI ACCESS

DYAH LISTYOWATI

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

PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026

Tim Penguji pada Ujian Skripsi:

1. Akbar Rizki, S. Stat., M.Si

Judul Skripsi : Implementasi Model IndoBERT Untuk Menangani Ketidakseimbangan Data Dalam Analisis Sentimen Ulasan Aplikasi KAI Access
Nama : Dyah Listyowati
NIM : G14190061

Disetujui oleh

Pembimbing 1:
Dr. Anwar Fitrianto, S.Si., M.Sc

Pembimbing 2:
Ir. Aam Alamudi, M.Si

Diketahui oleh

Ketua Program Studi:
Dr. Bagus Sartono, S.Si., M.Si.
NIP. 19780411 200501 1002

Tanggal Ujian:
6 Februari 2026

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 yang dilaksanakan sejak bulan Februari 2023 sampai bulan Januari 2026 ini ialah analisis sentimen ulasan aplikasi dengan judul "Implementasi Model IndoBERT Untuk Menangani Ketidakseimbangan Data Dalam Analisis Sentimen Ulasan Aplikasi KAI Access".

Terima kasih penulis ucapkan kepada seluruh pihak yang telah berperan sejak penulis menempuh studi sampai proses penulisan karya ilmiah, diantaranya

1. Ayah, Ibu, dan Kakak selaku orang tua dan saudara yang telah memberi doa dan dukungan,
2. Dr. Anwar Fitrianto, S.Si., M.Sc dan Ir. Aam Alamudi, M.Si, selaku Dosen Pembimbing Skripsi yang telah memberikan arahan dalam penulisan karya ilmiah ini,
3. Akbar Rizki, S. Stat., M.Si selaku Dosen Penguji Skripsi yang telah membantu saya dalam mengevaluasi dalam penulisan karya ilmiah ini,
4. Seluruh teman-teman Statistika IPB Angkatan 56
5. Seluruh pihak terlibat yang telah membantu penyelesaian studi yang tidak dapat saya sebutkan satu persatu

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan. Penulis menyadari bahwa karya ilmiah ini masih belum sempurna. Penulis memohon maaf atas kesalahan dan kekurangan dalam karya ilmiah ini. Penulis menyambut dengan baik apabila ada saran dan kritik terkait karya ilmiah ini.

Bogor, April 2026

Dyah Listyowati

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR