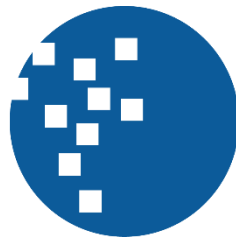


**ANALISIS SENTIMEN PENGALAMAN NASABAH PADA
KCU/KCP BERBASIS TF-IDF NAIVE BAYES DAN
INDOBERT DI PT BANK XYZ**



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

Laporan Skripsi

Edmund Daniel Valentino

00000068115

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

**ANALISIS SENTIMEN PENGALAMAN NASABAH PADA
KCU/KCP BERBASIS TF-IDF NAIVE BAYES DAN
INDOBERT DI PT BANK XYZ**



Laporan Skripsi

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Komputer

Edmund Daniel Valentino

00000068115

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Edmund Daniel Valentino

Nomor Induk Mahasiswa : 00000068115

Program Studi : Sistem Informasi

Skripsi dengan judul:

ANALISIS SENTIMEN PENGALAMAN NASABAH PADA KCU/KCP
BERBASIS TF-IDF NAIVE BAYES DAN INDOBERT PT BANK XYZ

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 28 November 2025

UNIVERS
MULTIM
NUSANT



(Edmund Daniel Valentino)

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Edmund Daniel Valentino

NIM : 00000068115

Program Studi : Sistem Informasi

Jenjang : S1

Judul Karya Ilmiah : ANALISIS SENTIMEN PENGALAMAN NASABAH PADA KCU/KCP BERBASIS TF-IDF NAIVE BAYES DAN INDOBERT DI PT BANK XYZ

Menyatakan dengan sesungguhnya bahwa saya bersedia* (pilih salah satu):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☒ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 28 November 2025



(Edmund Daniel Valentino)

* Pilih salah satu

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

(Kata Pengantar dapat dikembangkan dan harus meliputi ucapan rasa syukur, tujuan pembuatan tugas akhir, ucapan terima kasih, dan harapan pada hasil Tugas Akhir ini.)

Mengucapkan terima kasih

1. Dr. Ir. Andrey Andoko, M.Sc. Ph.D., selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti, S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ibu Monika Evelin Johan, S.Kom., M.M.S.I., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Ibu Fanny Suhendra, sebagai Pembimbing di tempat kerja magang yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
6. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat menjadi referensi yang bermanfaat dan memberikan dampak positif bagi para pembaca.

Tangerang, 28 November 2025



(Edmund Daniel Valentino)

ANALISIS SENTIMEN PENGALAMAN NASABAH PADA KCU/KCP BERBASIS TF-IDF NAIVE BAYES DAN INDOBERT DI PT BANK XYZ

Edmund Daniel Valentino

ABSTRAK

Penelitian ini berfokus pada penerapan analisis sentimen terhadap *customer experience* nasabah di Kantor Cabang Utama (KCU) dan Kantor Cabang Pembantu (KCP) Bank XYZ. Dataset diperoleh dari hasil survei kepuasan nasabah yang terlebih dahulu melalui proses pelabelan manual untuk menentukan kelas sentimen positif, netral, dan negatif. Penelitian dilaksanakan menggunakan metodologi CRISP-DM yang mencakup tahapan *business understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Dua pendekatan model digunakan dan dibandingkan, yaitu model klasik TF-IDF dengan Naive Bayes menggunakan GridSearchCV serta model *transformer* IndoBERT dengan Bayesian Optimization. Hasil pengujian menunjukkan bahwa model TF-IDF Naive Bayes mencapai akurasi sebesar 87%, sedangkan model IndoBERT menghasilkan performa yang lebih unggul dengan akurasi sebesar 96%. Meski demikian, model masih menghadapi tantangan pada kelas sentimen netral akibat ketidakseimbangan data. Model terbaik kemudian diimplementasikan dalam bentuk aplikasi berbasis Streamlit yang mampu mengolah file Excel berisi komentar nasabah dan menampilkan hasil klasifikasi sentimen secara interaktif. Penelitian ini membuktikan bahwa IndoBERT efektif untuk analisis sentimen berbahasa Indonesia dan dapat mendukung evaluasi kualitas layanan perbankan secara lebih objektif dan berbasis data.

Kata kunci: Analisis Sentimen, Bayesian Optimization, CRISP-DM, Customer Experience, GridSearchCV, IndoBERT, Naive Bayes, Streamlit

SENTIMENT ANALYSIS OF CUSTOMER EXPERIENCE AT KCU/KCP USING TF-IDF NAIVE BAYES AND INDOBERT IN PT BANK XYZ

Edmund Daniel Valentino

ABSTRACT (English)

This research focuses on the application of sentiment analysis to customer experience data from customers of the Main Branch Office (KCU) and Sub-Branch Office (KCP) of Bank XYZ. The dataset was obtained from customer satisfaction surveys and first underwent a manual labeling process to determine positive, neutral, and negative sentiment classes. The study was conducted using the CRISP-DM methodology, which includes the stages of business understanding, data preparation, modeling, evaluation, and deployment. Two modeling approaches were implemented and compared: a traditional TF-IDF with Naive Bayes model using GridSearchCV, and a transformer-based IndoBERT model with Bayesian Optimization. The experimental results show that the TF-IDF Naive Bayes model achieved an accuracy of 87%, while the IndoBERT model demonstrated superior performance with an accuracy of 96%. However, the model still faces challenges in the neutral sentiment class due to data imbalance. The best-performing model was then deployed as a Streamlit-based application capable of processing Excel files containing customer comments and presenting interactive sentiment classification results. This study demonstrates that IndoBERT is effective for sentiment analysis in Indonesian language and can support a more objective and data-driven evaluation of banking service quality.

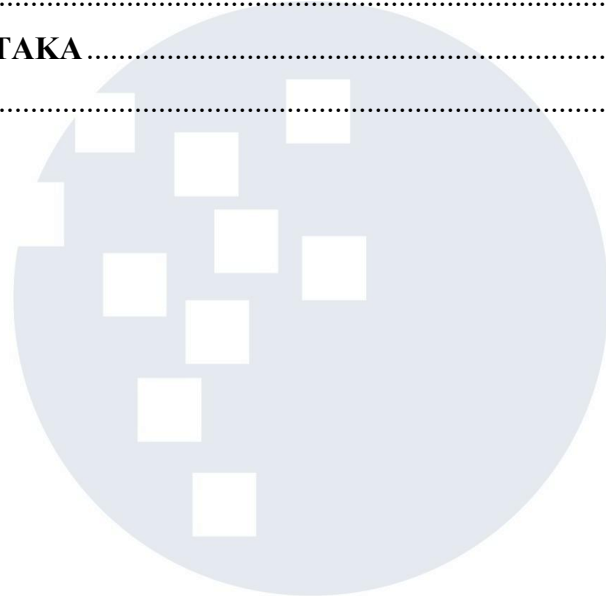
Keywords: *Bayesian Optimization, CRISP-DM, Customer Experience, GridSearchCV, IndoBERT, Naive Bayes, Sentiment Analysis, Streamlit*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
KATA PENGANTAR.....	v
ABSTRAK	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Batasan Masalah	5
1.4 Tujuan dan Manfaat Penelitian	6
1.4.1 Tujuan Penelitian	6
1.4.2 Manfaat Penelitian	6
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 Penelitian Terdahulu	9
2.2 Teori yang berkaitan	15
2.2.1 Analisis Sentimen	15
2.2.2 KCU dan KCP Bank.....	16
2.2.3 Customer Experience.....	17
2.2.4 Text Mining	18
2.2.5 Text Preprocessing	18
2.2.6 TF-IDF	19
2.3 Framework/Algoritma yang digunakan	20
2.3.1 CRISP-DM.....	20
2.3.2 Naive Bayes	23

2.3.3	BERT	24
2.3.4	IndoBERT	26
2.3.5	Manual Labeling	26
2.3.6	Hyperparameter Tuning	27
2.3.7	GridSearchCV	27
2.3.8	Bayesian Optimization	28
2.3.9	Confusion Matrix	28
2.3.10	Model Waterfall	30
2.3.11	Evaluasi Black Box	31
2.4	Tools/Software yang Digunakan	31
2.4.1	Python	31
2.4.2	Google Colab	32
2.4.3	Microsoft Excel	33
2.4.4	Streamlit	33
BAB III METODOLOGI PENELITIAN		34
3.1	Gambaran Umum Objek Penelitian	34
3.2	Metode Penelitian	34
3.2.1	Alur Penelitian	34
3.2.2	Metode Data Mining	36
3.3	Teknik Pengumpulan Data	50
3.4	Variabel Penelitian	50
3.5	Teknik Analisis Data	51
3.5.1	Bahasa Pemrograman yang Digunakan	51
3.5.2	Tools yang Digunakan	52
BAB IV ANALISIS DAN HASIL PENELITIAN		54
4.1	Implementasi Sistem	54
4.1.1	Business Understanding	54
4.1.2	Data Understanding	54
4.1.3	Data Preparation	60
4.1.4	Modeling (TF-IDF Naive Bayes)	71
4.1.5	Evaluation (TF-IDF Naive Bayes)	78
4.1.6	Modeling (IndoBERT)	83

4.1.7 Evaluation	101
4.1.8 Deployment	108
4.1.9 Feedback dan Evaluasi Aplikasi	125
BAB V SIMPULAN DAN SARAN	129
5.1 Simpulan	129
5.2 Saran	130
DAFTAR PUSTAKA	133
LAMPIRAN	143



UMN
 UNIVERSITAS
 MULTIMEDIA
 NUSANTARA

DAFTAR TABEL

Tabel 2.1 Tabel Jurnal Penelitian Terdahulu.....	9
Tabel 3.1 Perbandingan Metode Data Mining [72].....	37
Tabel 3.2 Gambaran Proses Manual Labeling	42
Tabel 3.3 Gambaran Proses Text Preprocessing Ringan.....	44
Tabel 3.4 Gambaran Proses Text Preprocessing Lanjutan.....	46
Tabel 3.5 Gambaran Hyperparameter Tuning Kedua Model.....	48
Tabel 3.6 Perbandingan Tools [74] [65] [75].....	52
Tabel 4.1 Dataset Survei Nasabah Customer Survey.....	55
Tabel 4.2 Contoh Data CUST_SURVEY_ID	56
Tabel 4.3 Contoh Data CUST_SURVEY_DT	57
Tabel 4.4 Contoh Data BRANCH_CD	57
Tabel 4.5 Contoh Data BRANCH_NM	58
Tabel 4.6 Contoh Data CITY	58
Tabel 4.7 Contoh Data CUST_FEEDBACK	59
Tabel 4.8 Manual Labeling oleh Dua Annotator.....	61
Tabel 4.9 Hasil Proses Text Preprocessing	65
Tabel 4.10 Hasil Proses Text Preprocessing Lanjutan.....	68
Tabel 4.11 Hasil Parameter dan CV Score Terbaik	76
Tabel 4.12 Hasil Classification Report TF-IDF Naive Bayes.....	79
Tabel 4.13 Hasil Confidence Level.....	82
Tabel 4.14 Parameter Penelitian Terdahulu	91
Tabel 4.15 Training Steps	94
Tabel 4.16 Training Steps Hyperparameter Tuning.....	99
Tabel 4.17 Classification Report	101
Tabel 4.18 Classification Report 2	104
Tabel 4.19 Skenario Pengujian Aplikasi Streamlit.....	126

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR GAMBAR

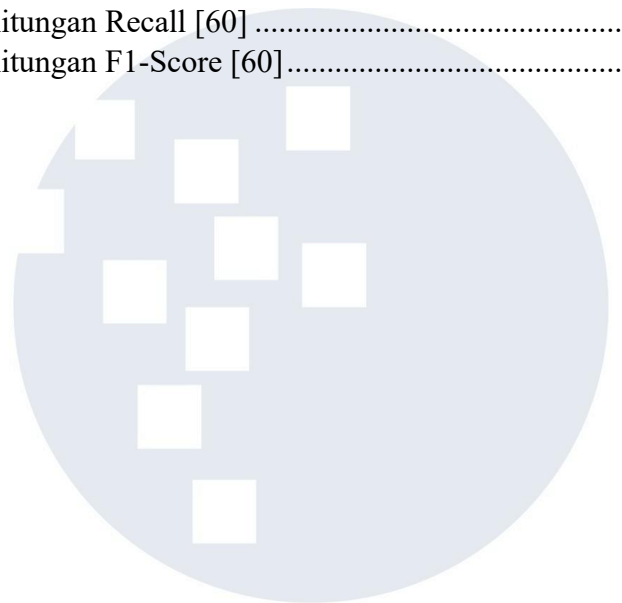
Gambar 2.1 Kantor Cabang Bank.....	16
Gambar 2.2 Metode CRISP-DM	21
Gambar 2.3 Arsitektur BERT.....	25
Gambar 2.4 Confusion Matrix.....	28
Gambar 2.5 Pemrograman Python	32
Gambar 2.6 Google Colab	32
Gambar 3.1 Diagram Alur Penelitian	35
Gambar 3.2 Metode Data Mining CRISP-DM.....	37
Gambar 3.3 Diagram Alur Data Understanding	40
Gambar 3.4 Diagram Alur Data Preparation	41
Gambar 3.5 Diagram Alur Modeling	47
Gambar 4.1 Distribusi Label Sentimen	62
Gambar 4.2 Library Text Preprocessing Ringan	63
Gambar 4.3 Function Text Preprocessing Ringan.....	64
Gambar 4.4 Penambahan Kolom CUST_EXP_CLEAN.....	65
Gambar 4.5 Inisialisasi NLTK dan Stemmer.....	67
Gambar 4.6 Function Text Preprocessing Lanjutan	68
Gambar 4.7 Kolom Baru CUST_EXP_CLEAN_STRING.....	69
Gambar 4.8 Label Encoding	70
Gambar 4.9 Pembagian Data	71
Gambar 4.10 Pipeline TF-IDF Naive Bayes	73
Gambar 4.11 Hyperparameter Tuning GridSearchCV	74
Gambar 4.12 Script Best Params dan Best CV Score.....	76
Gambar 4.13 Classification Report TF-IDF Naive Bayes.....	78
Gambar 4.14 Confusion Matrix TF-IDF Naive Bayes	80
Gambar 4.15 Script Confidence Level	81
Gambar 4.16 Data Splitting IndoBERT	84
Gambar 4.17 Load Dataset	85
Gambar 4.18 Inisialisasi Awal IndoBERT	86
Gambar 4.19 Inisialisasi Lanjutan IndoBERT	87
Gambar 4.20 Struktur Model di Python	88
Gambar 4.21 Function Metrik Evaluasi	90
Gambar 4.22 Konfigurasi Parameter IndoBERT	92
Gambar 4.23 Running Proses Training	93
Gambar 4.24 Prediksi Model.....	96
Gambar 4.25 Hyperparameter Tuning IndoBERT	97
Gambar 4.26 Parameter Terbaik.....	98
Gambar 4.27 Confusion Matrix dengan Sns	102
Gambar 4.28 Confusion Matrix.....	103
Gambar 4.29 Confusion Matrix Setelah Hyperparameter Tuning	105
Gambar 4.30 Train Vs Val Loss.....	107

Gambar 4.31 Menyimpan Training Model Terbaik	109
Gambar 4.32 Implementasi Aplikasi StreamLit (1)	110
Gambar 4.33 Implementasi Aplikasi StreamLit (2)	113
Gambar 4.34 Implementasi Aplikasi StreamLit (3)	115
Gambar 4.35 Membuat Link Streamlit melalui Ngrok.....	118
Gambar 4.36 Halaman Awal Streamlit.....	119
Gambar 4.37 Preview Dataset	120
Gambar 4.38 Hasil Analisis Sentimen.....	121
Gambar 4.39 Filter Berdasarkan Tanggal.....	122
Gambar 4.40 Searching Berdasarkan Branch Code	123
Gambar 4.41 Keyword Topik Negatif.....	124



DAFTAR RUMUS

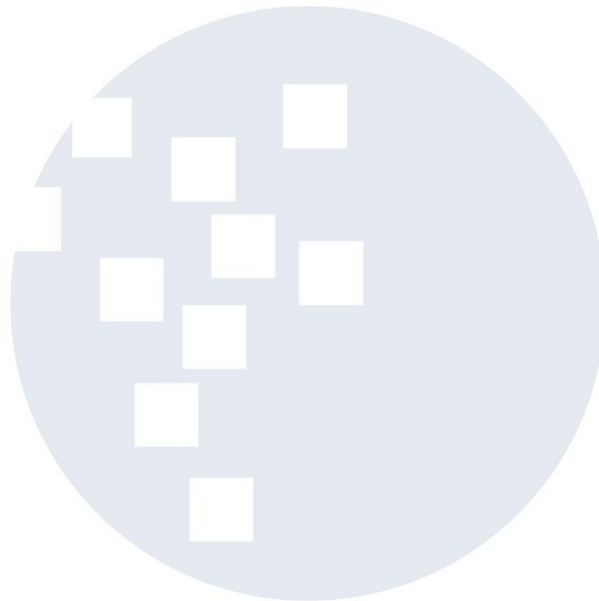
Rumus 2.1 Rumus Perhitungan TF [38]	20
Rumus 2.2 Rumus Perhitungan IDF [38].....	20
Rumus 2.3 Rumus Perhitungan Teorema Bayes [45]	24
Rumus 2.4 Perhitungan Accuracy [61]	29
Rumus 2.5 Perhitungan Precision [60].....	30
Rumus 2.6 Perhitungan Recall [60]	30
Rumus 2.7 Perhitungan F1-Score [60].....	30



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report	133
Lampiran B Form Konsultasi Bimbingan.....	134



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA