

**KOMPARASI METODE VEKTORISASI UNTUK
KLASIFIKASI AKUN BELANJA OPERASIONAL/NON
OPERASIONAL BERBASIS *NATURAL LANGUAGE
PROCESSING* (NLP)**

TESIS



Oleh:
GAHARA DIJERJA
NIM : 2311601849

**PROGRAM STUDI MAGISTER ILMU KOMPUTER
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

**JAKARTA
GASAL 2025/2026**

**KOMPARASI METODE VEKTORISASI UNTUK
KLASIFIKASI AKUN BELANJA OPERASIONAL/NON
OPERASIONAL BERBASIS *NATURAL LANGUAGE
PROCESSING* (NLP)**

Diajukan untuk memenuhi salah satu persyaratan memperoleh gelar
Magister Ilmu Komputer (M.Kom)



Oleh:
GAHARA DIJERJA
NIM : 2311601849

**PROGRAM STUDI MAGISTER ILMU KOMPUTER
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

**JAKARTA
GASAL 2025/2026**

ABSTRAK

Komparasi Metode Vektorisasi Untuk Klasifikasi Akun Belanja Operasional/Non Operasional Berbasis *Natural Language Processing* (NLP)

Oleh: Gahara Dijerja (NIM 2311601849)

Ketidakpatuhan dalam klasifikasi akun belanja, terutama pada akun belanja operasional lainnya (521119) dan belanja non-operasional lainnya (521219), berpotensi menurunkan kualitas penyajian laporan keuangan pemerintah karena kedua akun tersebut kerap dipakai sebagai akun penampungan sementara sehingga memunculkan risiko bias pencatatan. Penelitian ini bertujuan mengembangkan dan mengevaluasi sistem klasifikasi otomatis akun belanja pemerintah berbasis teks untuk membantu proses verifikasi dan meningkatkan konsistensi pembebanan akun. Data yang digunakan merupakan data riil transaksi SAKTI tahun 2024 yang merepresentasikan sembilan kelas akun belanja (521111–521219) dengan fokus pengujian pada pemisahan operasional dan non-operasional. Uraian transaksi diproses melalui tahapan text preprocessing (pembersihan teks, normalisasi, tokenisasi, penghapusan *stopword*, dan *stemming*), kemudian direpresentasikan menggunakan *TF-IDF*, *Bag-of-Words*, *Word2Vec*, serta pendekatan kontekstual berbasis *BERT*. Selanjutnya, vektor fitur diklasifikasikan menggunakan *Support Vector Machine (SVM)*, *eXtreme Gradient Boosting (XGBoost)*, dan *BERT*, dengan penanganan ketidakseimbangan kelas menggunakan *SMOTE* pada tahap pelatihan. Hasil evaluasi menunjukkan bahwa kombinasi *Word2Vec + XGBoost* memberikan performa terbaik dengan skor *cross-validation* tertinggi 84,50 sekaligus waktu inferensi tercepat 0,0048 detik. *BoW + XGBoost* mencatat *precision* tertinggi 86,21%, sedangkan *BERT* menghasilkan *confidence* hingga 99,89% namun memerlukan waktu komputasi lebih lama (0,1338 detik). Penelitian ini juga menghasilkan prototipe aplikasi berbasis web menggunakan Streamlit untuk mendukung inferensi dan evaluasi model secara langsung. Implikasi penelitian ini adalah tersedianya *decision support* yang dapat mengurangi penggunaan akun penampungan secara tidak tepat, mempercepat verifikasi transaksi, meningkatkan *audit-readiness*, serta menjadi dasar integrasi bertahap ke proses bisnis verifikasi dan pembaruan model berkala berbasis data transaksi terbaru.

Kata kunci: *BERT, Klasifikasi, Machine Learning, NLP, XGBoost.*

ABSTRACT

Comparison of Vectorization Methods for the Classification of Operational and Non-Operational Expenditure Accounts Based on Natural Language Processing (NLP)

By: Gahara Dijerja (NIM 2311601849)

Non-compliance in expenditure account classification, particularly for “Other Operating Expenditures” (521119) and “Other Non-Operating Expenditures” (521219), can reduce the quality of government financial reporting because both accounts are often used as temporary holding accounts, thereby increasing the risk of recording bias. This study aims to develop and evaluate an automatic text-based classification system for government expenditure accounts to support the verification process and improve the consistency of account charging. The study uses real transaction data from the SAKTI system in 2024, representing nine expenditure account classes (521111–521219), with an emphasis on distinguishing operating versus non-operating transactions. Transaction descriptions are processed through standard text preprocessing (text cleaning, normalization, tokenization, stopword removal, and stemming), then transformed into numerical representations using TF-IDF, Bag-of-Words, Word2Vec, and a contextual approach based on BERT. The resulting feature vectors are classified using Support Vector Machine (SVM), eXtreme Gradient Boosting (XGBoost), and BERT, while class imbalance is addressed using SMOTE during the training stage. The evaluation results indicate that the Word2Vec + XGBoost combination achieves the best overall performance, obtaining the highest cross-validation score of 84.50 and the fastest inference time of 0.0048 seconds. Although BoW + XGBoost records the highest precision (86.21%), Word2Vec + XGBoost is recommended due to its semantic stability. In contrast, BERT yields the highest prediction confidence (up to 99.89%) but requires the longest computation time (0.1338 seconds). This study also delivers a web-based prototype application developed with Streamlit to enable direct model inference and evaluation. The practical implication of this research is the availability of a decision-support mechanism that can reduce inappropriate use of holding accounts, accelerate transaction verification, enhance audit readiness, and serve as a foundation for gradual integration into verification workflows and periodic model updates using the most recent transaction data.

Keywords: *BERT, Classification, Machine Learning, NLP, XGBoost.*



BERITA ACARA SIDANG PENDADARAN TESIS

S/UBL/PPS/201/0679/I/26

Pada hari ini, Senin 19 Januari 2026 telah dilaksanakan Ujian Sidang Tesis sebagai berikut:

Judul: KOMPARASI METODE VEKTORISASI UNTUK KLASIFIKASI AKUN BELANJA OPERASIONAL/NON OPERASIONAL BERBASIS NATURAL LANGUAGE PROCESSING (NLP)

Nama : Gahara Dijerja
NIM : 2311601849
Dosen Pembimbing : Dwi Pebrianti, S.T, M.Eng, Ph.D.

Berdasarkan Permasalahan, Metodologi, Tinjauan Pustaka, Tatatulis, dan Presentasi, Maka Mahasiswa tersebut dinyatakan :

LULUS

dengan nilai angka : **90** huruf : **A**

Mahasiswa tersebut di atas wajib menyerahkan hasil perbaikan tulisan Tesis dalam bentuk terjilid sesuai dengan Panduan Perbaikan Tesis, selambat-lambatnya Rabu 18 Februari 2026.

Panitia Penguji:

- 1 Ketua Dr. Ir. Achmad Solichin, S.Kom., M.T.I.
- 2 Anggota Dr. Samidi, S.Kom., M.M., M.Kom.
- 3 Moderator Dwi Pebrianti, S.T, M.Eng, Ph.D.

Keterangan:

Nilai Huruf: A:85-100 A-:80-84,99 B+:75-79,99 B:70-74,99 B-:65-69,99 C:60-64,99 D:40-59,99 E-:0-39,99



LEMBAR PENGESAHAN

Nama : Gahara Dijerja
Nomor Induk Mahasiswa : 2311601849
Program Studi : Magister Ilmu Komputer
Bidang Peminatan : Teknologi Sistem Informasi
Jenjang Studi : Strata 2
Judul : KOMPARASI METODE VEKTORISASI UNTUK KLASIFIKASI AKUN
BELANJA OPERASIONAL/NON OPERASIONAL BERBASIS NATURAL
LANGUAGE
PROCESSING (NLP)



Laporan Tugas Akhir ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Senin 19 Januari 2026

Tim Penguji:

Ketua : Dr. Ir. Achmad Solichin, S.Kom., M.T.I.
Anggota : Dr. Samidi, S.Kom., M.M., M.Kom.
Pembimbing : Dwi Pebrianti, S.T, M.Eng, Ph.D.
Ketua Program Studi : Dr. Rusdah, S.Kom., M.Kom.



KEPUTUSAN DEKAN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR

NOMOR : K/UBL/FTI/000/002/11/25

**TENTANG :
PENGANGKATAN DOSEN PENGUJI TESIS
PROGRAM STUDI MAGISTER ILMU KOMPUTER
SEMESTER GASAL TAHUN AKADEMIK 2025/2026**

DEKAN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR

- Menimbang : 1) Bahwa untuk meningkatkan kualitas bagi mahasiswa Program Studi Magister Ilmu Komputer Fakultas Teknologi Informasi Universitas Budi Luhur Jenjang Studi Strata Dua (S-2) dalam proses tesis, perlu adanya penguji yang sesuai dengan materi tesis mahasiswa;
- 2) Bahwa dalam hubungan itu dipandang perlu mengangkat dosen–dosen yang tepat untuk menguji mahasiswa Program Studi Magister Ilmu Komputer Fakultas Teknologi Informasi Universitas Budi Luhur dalam sidang tesis;
- Mengingat : 1) Undang – undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan Tinggi;
- 2) Undang – undang Republik Indonesia Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional;
- 3) Peraturan Pemerintah Republik Indonesia Nomor 66 Tahun 2010 tentang Pengelolaan dan Penyelenggaraan Pendidikan;
- 4) Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi (Permendikbudristek) Nomor 39 Tahun 2025 tentang Penjaminan Mutu Pendidikan Tinggi;
- 5) Keputusan Ketua Yayasan Pendidikan Budi Luhur Cakti Nomor: K/YBLC/KEP/000/216/06/2023 tentang Statuta Universitas Budi Luhur;
- 6) Surat Keputusan Yayasan Pendidikan Budi Luhur Cakti Surat Keputusan Yayasan Pendidikan Budi Luhur Cakti No: K/YBLC/KEP/000/384/09/25 tanggal 22 September 2025 tentang Pembeastugasan Dan Pengangkatan Kembali Pejabat Struktural Universitas Budi Luhur Periode 2024-2028.

MEMUTUSKAN

- Menetapkan :
PERTAMA : Membentuk Dosen Penguji Tesis mahasiswa Program Studi Magister Ilmu Komputer jenjang S-2 Fakultas Teknologi Informasi Universitas Budi Luhur periode Semester Gasal 2025/2026;
- KEDUA : Menunjuk dan mengangkat Pejabat dan Dosen Program Studi Magister Ilmu Komputer Fakultas Teknologi Informasi Universitas Budi Luhur, yang nama-namanya tercantum dalam lampiran keputusan ini sebagai anggota yang dimaksud dalam diktum pertama;
- KETIGA : Tugas pokok dosen penguji tesis adalah menguji dan memberikan penilaian kepada mahasiswa Program Studi Magister Ilmu Komputer Fakultas Teknologi Informasi Universitas Budi Luhur dalam ujian sidang tesis;
- KEEMPAT : Dalam melaksanakan tugasnya Dosen penguji tesis bertanggung jawab kepada Dekan Fakultas Teknologi Informasi Universitas Budi Luhur melalui Ketua Program Studi.
- KELIMA : Keputusan ini berlaku sejak tanggal ditetapkan, sampai dengan akhir Semester Gasal Tahun Akademik 2025/2026 dan akan diunah sebagaimana mestinya apabila dikemudian hari terdapat kekeliruan.

Ditetapkan di : Jakarta

Pada Tanggal : 12 November 2025

=====
Dekan Fakultas Teknologi Informasi



Dr. Ir. Achmad Solichin, S.Kom., M.T.I



LAMPIRAN KEPUTUSAN DEKAN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR

NOMOR : K/UBL/FTI/000/002/11/25

NAMA-NAMA DOSEN PENGUJI TESIS PROGRAM STUDI MAGISTER ILMU KOMPUTER SEMESTER GASAL TAHUN AKADEMIK 2025/2026

NO	NIP	NAMA DOSEN
1	180057	Arif Bramantoro., Ph.D
2	230006	Dr. Akhmad Unggul Priantoro
3	030574	Dr. Anton Satria Prabuwono, ST., S.Si., M.M
4	170029	Dr. Darmawan Baginda Napitupulu, ST., M.Kom
5	030557	Dr. Imelda, S.Kom., M.Kom
6	080062	Dr. Indra, S.Kom., M.T.I
7	050023	Dr. Ir. Achmad Solichin, S.Kom., M.T.I
8	960012	Dr. Ir. Deni Mahdiana, S.Kom., M.M., M.Kom
9	970020	Dr. Ir. Gandung Triyono, M.Kom
10	930008	Dr. Ir. Goenawan Brotosaputro, S.Kom., M.Sc
11	980002	Dr. Ir. Jan Everhard Riwurohi, M.T
12	920024	Dr. Ir. Mardi Hardjianto, S.Kom., M.Kom
13	870021	Dr. Ir. Nazori AZ, M.T
14	960010	Dr. Ir. Utomo Budiyanto, M.Kom., M.Sc
15	890001	Dr. Krisna Adiyarta Musodo
16	960008	Dr. Muhammad Syafrullah, M.Kom., M.Sc
17	080110	Dr. Rudi Rusdiah, M.B.A, M.A
18	010016	Dr. Rusdah, S.Kom., M.Kom
19	870021	Dr. Indra Riyanto, S.T., M.T
20	100070	Dr. Samidi., S.Kom., M.M., M.Kom
21	170037	Dwi Febrianti, Ph.D
22	160001	Indra Nugraha Abdullah, S.Kom., M.Sc., Ph.D
23	150064	Ir. Denni Kurniawan, ST., M.IT., Ph.D
24	170032	Luhur Bayuaji, ST., M.Eng., Ph.D
25	020004	Prof. Dr. Ir. Arief Wibowo, S.Kom., M.Kom
26	930011	Prof. Dr. Ir. Hari Soetanto, S.Kom., M.Sc
27	850005	Prof. Dr. Ir. Wendi Usino, M.M., M.Sc.
28	830009	Prof. Dr. Moedjiono, M.Sc
29	070009	Prof. Ir. Dana Indra Sensuse, Ph.D
30	120055	Setyawan Widyarto, M.Sc., Ph.D

Ditetapkan di : Jakarta

Pada Tanggal : 12 November 2025

=====
Dekan Fakultas Teknologi Informasi



Dr. Ir. Achmad Solichin, S.Kom., M.T.I