SKRIPSI

KLASIFIKASI SERTIFIKAT BERDASARKAN MATA KULIAH DALAM PROGRAM REKOGNISI PEMBELAJARAN LAMPAU BERBASIS NLP MENGGUNAKAN BERT

DIMAS SAPUTRA
NPM 21081010151

DOSEN PEMBIMBING
Dr. Ir. I Gede Susrama Mas Diyasa, ST. MT. IPU
Eva Yulia Puspaningrum, S.Kom., M.Kom

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025

LEMBAR PENGESAHAN

KLASIFIKASI SERTIFIKAT BERDASARKAN MATA KULIAH DALAM PROGRAM REKOGNISI PEMBELAJARAN LAMPAU BERBASIS NLP MENGGUNAKAN BERT

Oleh :
DIMAS SAPUTRA
NPM. 21081010151

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakulktas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 27 Mei 2025

Menyetujui

Dr. Ir. I Gede Susrama Mas Diyasa, ST. MT. IPU NIP. 19700619 202121 1 009	.....	(Pembimbing I)
Eva Yulia Puspaningrum, S.Kom., M.Kom NIP. 19890705 202121 2 002	.....	(Pembimbing II)
Made Hanindia Prami Swari, S.Kom, M.Cs NIP. 19890205 201803 2 001	.....	(Ketua Penguji)
Achmad Junaidi, S.Kom., M.Kom NPT. 3 7811 04 0199 1	.....	(Anggota Penguji)

Mengetahui,
Dekan Fakultas Ilmu Komputer

Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

KLASIFIKASI SERTIFIKAT BERDASARKAN MATA KULIAH DALAM PROGRAM REKOGNISI PEMBELAJARAN LAMPAU BERBASIS NLP MENGGUNAKAN BERT

Oleh :
DIMAS SAPUTRA
NPM. 21081010151

Menyetujui,

Kooridnator Program Studi Informatika

Fakultas Ilmu Komputer

Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP. 19820211 202121 2 005

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama	: Dimas Saputra
NPM	: 21081010151
Program	: Sarjana (S1)
Program Studi	: Informatika
Fakultas	: Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 27 Mei 2025
Yang Membuat Pernyataan,

Dimas Saputra
NPM. 21081010151

ABSTRAK

Nama Mahasiswa / NPM	:	Dimas Saputra / 21081010151
Judul Skripsi	:	Klasifikasi Sertifikat Berdasarkan Mata Kuliah Dalam Program Rekognisi Pembelajaran Lampau Berbasis NLP Menggunakan BERT
Dosen Pembimbing	:	1. Dr. Ir. I Gede Susrama Mas Diyasa, ST. MT. IPU 2. Eva Yulia Puspaningrum, S.Kom., M.Kom

Perkembangan teknologi informasi yang pesat mendorong kebutuhan akan sistem otomatisasi dalam pengakuan pembelajaran non-formal, salah satunya melalui Rekognisi Pembelajaran Lampau (RPL). Sertifikat kompetensi yang diperoleh dari pelatihan mandiri sering kali tidak terintegrasi secara sistematis ke dalam program akademik. Penelitian ini bertujuan untuk mengembangkan model klasifikasi otomatis yang dapat mengelompokkan sertifikat ke dalam mata kuliah terkait dalam lingkup program studi Informatika. Tujuh kategori mata kuliah ditentukan sebagai kelas target, meliputi: Machine Learning, Pemrograman Web, Desain Antarmuka, Jaringan Komputer, Aplikasi Game, Pemrograman Mobile, dan Manajemen Proyek. Metodologi yang digunakan meliputi ekstraksi data sertifikat berbentuk PDF melalui proses OCR menggunakan PyTesseract, diikuti dengan tahap praproses teks dan data *Augmentation* untuk memperbaiki distribusi kelas. Model yang diimplementasikan adalah BERT (*Bidirectional Encoder Representations from Transformers*) dengan dua konfigurasi, yakni bert-base-uncased dan bert-base-multilingual-uncased. Lima skenario augmentasi diuji, yaitu tanpa augmentasi, *Character Insertion*, *Character Deletion*, *Back Translation*, dan *Synonym Replacement*. Hasil evaluasi menunjukkan bahwa konfigurasi bert-base-uncased dengan teknik augmentasi *Synonym Replacement* menghasilkan performa terbaik, dengan akurasi validasi mencapai 95,54% dan *F1-score* sebesar 0,97. Temuan ini menegaskan efektivitas BERT dalam klasifikasi teks berbahasa Indonesia-Inggris dan manfaat augmentasi semantik dalam meningkatkan generalisasi model. Sebagai implementasi praktis, penelitian ini juga mengembangkan prototipe layanan berbasis Streamlit yang memungkinkan klasifikasi otomatis sertifikat melalui antarmuka sederhana. Model dan sistem yang dikembangkan diharapkan dapat mendukung integrasi kompetensi non-formal ke dalam kurikulum akademik secara lebih efisien dan akurat.

Kata kunci : Klasifikasi Teks, BERT, Rekognisi Pembelajaran Lampau, Augmentasi Data, Pemrosesan Bahasa Alami

ABSTRACT

Student Name / NPM	:	Dimas Saputra / 21081010151
Thesis Title	:	Classification of Certificates Based on Courses in Prior Learning Recognition Program Using NLP with BERT
Advisor	:	1. Dr. Ir. I Gede Susrama Mas Diyasa, ST. MT. IPU 2. Eva Yulia Puspaningrum, S.Kom., M.Kom

ABSTRACT

The rapid advancement of information technology has increased the demand for automated systems to recognize non-formal learning, particularly through the Recognition of Prior Learning (RPL) program. Competency certificates obtained from independent training programs are often not systematically integrated into academic curricula. This study aims to develop an automated classification model capable of categorizing certificates into relevant course subjects within the Informatics study program. Seven target course categories were defined, including Machine Learning, Web Programming, Interface Design, Computer Networks, Game Applications, Mobile Programming, and Project Management. The methodology includes extracting text from certificate documents in PDF format through Optical Character Recognition (OCR) using PyTesseract, followed by text praproses and data Augmentation to improve class distribution. The implemented model is BERT (Bidirectional Encoder Representations from Transformers), evaluated in two configurations: bert-base-uncased and bert-base-multilingual-uncased. Five data scenarios were tested: no Augmentation, Character Insertion, Character Deletion, Back Translation, and Synonym Replacement. Evaluation results indicate that the bert-base-uncased configuration with Synonym Replacement Augmentation yielded the best performance, achieving a validation accuracy of 95.54% and an *F1-score* of 0.97. These findings confirm the effectiveness of BERT for text classification in both Indonesian and English, and highlight the benefit of semantic-based Augmentation techniques in improving model generalization. As a practical implementation, this research also developed a prototype service using the Streamlit framework, enabling automated certificate classification through a user-friendly interface. The model and system developed are expected to support the efficient and accurate integration of non-formal competencies into academic programs.

Keywords: Text Classification, BERT, Recognition of Prior Learning, Data Augmentation, Natural Language Processing

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **"Klasifikasi Sertifikat Berdasarkan Mata Kuliah Dalam Program Rekognisi Pembelajaran Lampau Berbasis NLP Menggunakan BERT"** dapat terselesaikan dengan baik.

Peneliti menyadari tidak akan dapat menyelesaikan skripsi ini dengan baik tanpa rida, bimbingan, saran, motivasi, dan bantuan dari berbagai pihak. Pada kesempatan ini, penulis mengucapkan terimakasih yang sebesar–besarnya kepada:

1. Kedua orang tua serta kakak, selaku keluarga yang selalu memberikan doa, dukungan moril maupun materiil, serta dukungan yang tiada henti.
2. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT. selaku Rektor Universitas Pembangunan Nasional "Veteran" Jawa Timur.
3. Ibu Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakulas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur.
4. Ibu Fetty Try Anggraeny, S.Kom, M.Kom. Selaku Koordinator Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur.
5. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom. selaku Koordinator Skripsi, yang telah membantu dalam proses administrasi dan pelaksanaan skripsi.
6. Bapak Dr. Ir. I Gede Susrama Mas Diyasa, ST., MT., IPU. selaku Dosen Pembimbing pertama, yang telah sabar memberikan bimbingan, motivasi, dan waktunya dalam proses bimbingan penyusunan skripsi
7. Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom. selaku Dosen Pembimbing kedua, yang telah memberikan arahan dan saran dalam proses penyusunan skripsi.
8. Ibu Made Hanindia Prami Swari, S.Kom., M.Cs., selaku Ketua Penguji dan Dosen Wali, yang telah memberikan bimbingan, nasehat, kritik dan saran yang sangat berharga selama masa perkuliahan dan proses penyusunan skripsi ini.

9. Bapak Achmad Junaidi, S.Kom., M.Kom., selaku Anggota Penguji, yang telah memberikan kritik dan masukan yang sangat bermanfaat dalam proses penyusunan skripsi ini.
10. Seluruh staf Tata Usaha dan civitas akademika Universitas Pembangunan Nasional "Veteran" Jawa Timur atas bantuan administrasi dan pelayanan selama penulis menempuh pendidikan.
11. Seluruh keluarga besar dan kerabat yang senantiasa memberikan doa dan dukungan.
12. Teman-teman seperjuangan yang selalu memberikan semangat, motivasi, dan kebersamaan selama proses penyusunan skripsi.
13. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu secara langsung maupun tidak langsung dalam penyusunan skripsi ini.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 27 Mei 2025

Dimas Saputra

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR