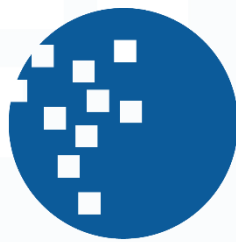


PERANCANGAN *EXTREME MULTI-CLASS CLASSIFIER*

UNTUK SEGMENTASI PRODUK PADA

KAWAN LAMA GROUP



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

LAPORAN MBKM

Vincencius Christiano Tjokro

0000053887

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK & INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

PERANCANGAN *EXTREME MULTI-CLASS CLASSIFIER*

UNTUK SEGMENTASI PRODUK PADA

KAWAN LAMA GROUP



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

LAPORAN MBKM

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer (S.Kom)

Vincencius Christiano Tjokro

0000053887

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK & INFORMATIKA

UNIVERSITAS MULTIMEDIA NUSANTARA

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Vincencius Christiano Tjokro

Nomor Induk Mahasiswa : 00000053887

Program studi : Sistem Informasi

Laporan MBKM Penelitian dengan judul:

“Perancangan *Extreme Multi-Class Classifier* Untuk Segmentasi Produk Pada Kawan Lama Group”

merupakan hasil karya saya sendiri bukan plagiat dari karya 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 MBKM, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk laporan MBKM yang telah saya tempuh.

Tangerang, 20 Mei 2024



Vincencius Christiano Tjokro

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertandatangan di bawah ini:

Nama : Vincencius Christiano Tjokro

NIM : 00000053887

Program Studi : Sistem Informasi.

Fakultas : Fakultas Teknik & Informatika.

Jenis Karya : Laporan MBKM

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Noneklusif (Non-exclusive Royalty-Free Right) atas karya ilmiah saya yang berjudul:

PERANCANGAN EXTREME MULTI-CLASS CLASSIFIER UNTUK SEGMENTASI PRODUK PADA KAWAN LAMA GROUP

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 20 Mei 2024

Yang menyatakan,



Vincencius Christiano Tjokro

KATA PENGANTAR

Puji Syukur atas selesainya penulisan laporan magang ini dengan judul: “PERANCANGAN *EXTREME MULTI-CLASS CLASSIFIER* UNTUK SEGMENTASI PRODUK PADA KAWAN LAMA GROUP” dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer (S.Kom) Jurusan Sistem Informasi Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan tugas akhir ini, sangatlah sulit bagi saya untuk menyelesaikan tugas akhir ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, M.A., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak 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 Dr. Irmawati, S.Kom., M.M.S.I., sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini.
5. Kepada Ibu Agustina Iftariani, Bapak Eduardus Gunarso, Bapak Albert Hadinata, serta seluruh pihak Kawan Lama Group yang mendukung selama proses magang dilakukan.
6. Keluarga yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan laporan MBKM ini.

Semoga karya ilmiah ini dapat memberikan dampak positif bagi para pembaca dan menjadi sumber informasi serta wawasan yang bermanfaat.

Tangerang, 20 Mei 2024



Vincencius Christiano Tjokro



PERANCANGAN *EXTREME MULTI-CLASS CLASSIFIER* UNTUK SEGMENTASI PRODUK PADA KAWAN LAMA GROUP

Vincencius Christiano Tjokro

ABSTRAK

Kawan Lama Group sebagai perusahaan konglomerasi memiliki banyak unit bisnis di bawahnya. Kompleksitas ini menghadirkan tantangan berupa integritas data produk. Untuk mengatasi permasalahan tersebut implementasi *Product Information Management* (PIM) dilakukan. Dalam rangka menunjang kesiapan data dalam implementasi sistem PIM, segmentasi data produk ke dalam taksonomi yang bersifat umum serta tidak terikat ke dalam unit bisnis manapun perlu dilakukan. Untuk meningkatkan proses segmentasi yang sebelumnya dilakukan secara manual, pengembangan model *Extreme Multi-Class Classifier* (XMC) dilakukan melalui *fine-tuning* pada model *deep learning* SBERT dengan *MultipleNegativeRanking* sebagai *loss function*. Proses pelatihan pertama menghasilkan model dengan akurasi sebesar 79.45%, penambahan data latih mampu meningkatkan akurasi sebesar 85.21%. Kehadiran model XMC secara keseluruhan mampu meningkatkan produktivitas segmentasi sebesar 80% dengan pendekatan *human-in-the-loop*, menggabungkan kemampuan model dalam segmentasi pada *threshold* yang telah ditentukan dengan manusia dalam proses validasi serta segmentasi secara manual untuk data produk yang gagal tersegmentasi. Dari 833 produk per orang per hari, hasil segmentasi meningkat menjadi 1500 produk per orang per hari.

Kata kunci: *Extreme Multi-Class Classifier*, Segmentasi Produk, SBERT

ERANCANGAN *EXTREME MULTI-CLASS CLASSIFIER* UNTUK SEGMENTASI PRODUK PADA KAWAN LAMA GROUP

Vincencius Christiano Tjokro

ABSTRACT (English)

Kawan Lama Group, as a conglomerate company, has numerous business units under its umbrella. This complexity presents a challenge in maintaining data integrity for products. To address this issue, the implementation of Product Information Management (PIM) is underway. In order to support data readiness for the PIM system implementation, product data segmentation into a generic taxonomy not tied to any specific business unit is necessary. To enhance the segmentation process, which was previously done manually, the development of an Extreme Multi-Class Classifier (XMC) model was carried out through fine-tuning on the SBERT deep learning model with MultipleNegativeRanking as the loss function. The initial training process resulted in a model with an accuracy of 79.45%, and the addition of training data increased the accuracy to 85.21%. The presence of the XMC model overall improved segmentation productivity by 80% using a human-in-the-loop approach, combining the model's segmentation capabilities at predefined thresholds with human validation and manual segmentation for products that fail to segment. From 833 products per person per day, the segmentation output increased to 1500 products per person per day.

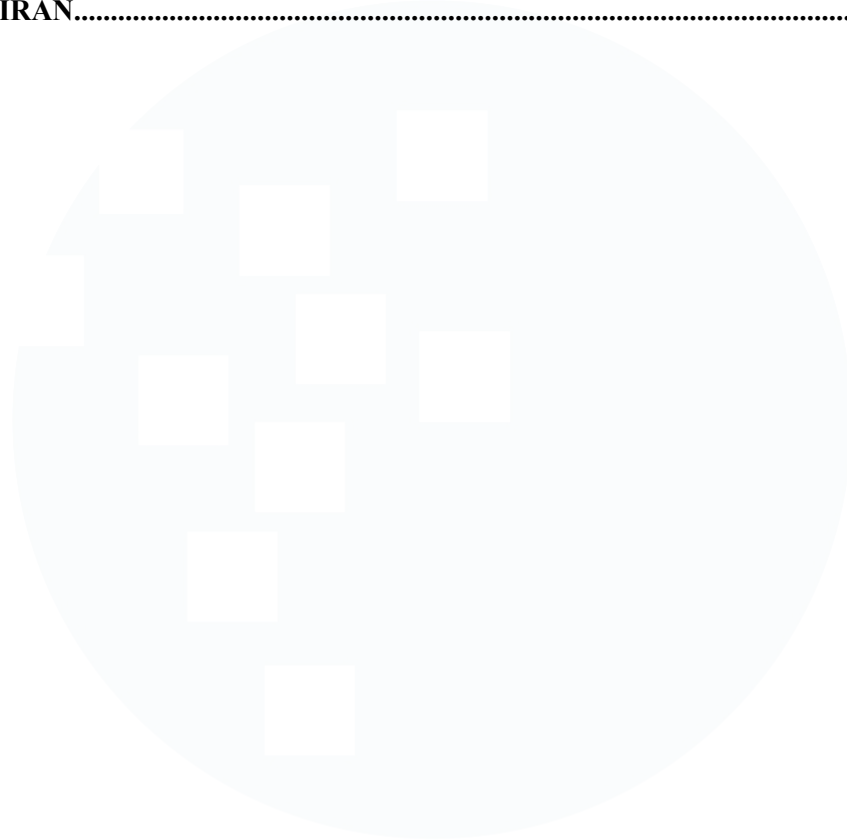
Keywords: *Extreme Multi-Class Classifier, product segmentation, SBERT*

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	I
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	II
KATA PENGANTAR	III
ABSTRAK	V
ABSTRACT (English)	VI
DAFTAR ISI	VII
DAFTAR TABEL	IX
DAFTAR GAMBAR	X
DAFTAR LAMPIRAN	XII
BAB I	
PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Maksud dan Tujuan Kerja Magang.....	2
1.3. Waktu dan Prosedur Pelaksanaan Kerja Magang.....	3
1.3.1. Waktu Pelaksanaan.....	3
BAB II	
GAMBARAN UMUM PERUSAHAAN	6
2.1 Sejarah Singkat Perusahaan.....	6
2.1.1 Visi Misi.....	7
2.1.2 Nilai-Nilai Perusahaan.....	7
2.2 Struktur Organisasi Perusahaan.....	8
BAB III	
PELAKSANAAN KERJA MAGANG	10
3.1 Kedudukan dan Koordinasi.....	10
3.2 Tugas dan Uraian Kerja Magang.....	11
3.2.1 Pengenalan Perusahaan dan Proyek.....	13
3.2.2 Segmentasi Data secara Manual.....	13
3.2.3 Research Metode Machine Learning.....	16
3.2.4 Pengembangan Model XMC.....	19
3.2.5 Pengembangan Antarmuka Human-in-the-Loop.....	49
3.2.6 Segmentasi Data dengan pendekatan Human-in-the-Loop.....	57
3.2.7 Validasi Artikel pasca Segmentasi.....	59
3.3 Kendala yang Ditemukan.....	61
3.4 Solusi atas Kendala yang Ditemukan.....	62
BAB IV	
SIMPULAN DAN SARAN	64
4.1 Simpulan.....	64
4.2 Saran.....	65

VII

DAFTAR PUSTAKA.....	67
LAMPIRAN.....	68



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 1.1. Timeline Kegiatan Magang.....	3
Tabel 3.1. Penjabaran Pekerja per Minggu.....	12
Tabel 3.2. Hasil Initial Testing.....	22
Tabel 3.3. Hasil Fine-tuning NMR-V0.1 per Epoch.....	23
Tabel 3.4. Perbandingan Akurasi Model Pre-trained.....	48
Tabel 3.5. Hasil Fine-tuning CSE-NMR-V.1 per Epoch.....	48
Tabel 3.6. Perbandingan Proses Manual dengan Human-in-the-loop.....	58



DAFTAR GAMBAR

Gambar 2.1. Logo Kawan Lama Group.....	6
Gambar 2.2. Struktur Organisasi Master Data Management.....	9
Gambar 3.1. Diagram Alur Kerja Divisi.....	11
Gambar 3.2. Diagram Segmentasi Manual.....	14
Gambar 3.3. Ilustrasi Struktur Data Produk.....	16
Gambar 3.4. Arsitektur SBERT Pengukuran Similaritas.....	18
Gambar 3.5. Flowchart Fine-tuning Model SBERT.....	20
Gambar 3.6. Struktur Folder ‘NMR-V0.1’.....	24
Gambar 3.7. Flowchart Proses Segmentasi.....	25
Gambar 3.8. Gambar Tabel Hasil Prediksi dengan ID.....	26
Gambar 3.9. Fungsi cols_exist() pada LAMHelper.....	27
Gambar 3.10. Fungsi get_rows_missing_values() pada LAMHelper.....	28
Gambar 3.11. Fungsi join_no_duplicate() pada LAMHelper.....	29
Gambar 3.12. Fungsi get_unique_row() pada LAMHelper.....	29
Gambar 3.13. Fungsi split_data() pada LAMHelper.....	30
Gambar 3.14. Fungsi merge_true_predict() pada LAMHelper.....	30
Gambar 3.15. Fungsi get_taxonomy_cols() pada LAMHelper.....	31
Gambar 3.16. Fungsi get_taxonomy_merged() pada LAMHelper.....	31
Gambar 3.17. Fungsi get_negative_pair() pada LAMHelper.....	32
Gambar 3.18. Fungsi get_key_col() pada LAMHelper.....	32
Gambar 3.19. Fungsi get_subset_cols() pada LAMHelper.....	33
Gambar 3.20. Fungsi get_article_merged() pada LAMHelper.....	34
Gambar 3.21. Fungsi __init__() pada SBERT_MNR.....	35
Gambar 3.22. Fungsi enkapsulasi pada SBERT_MNR.....	36
Gambar 3.23. Fungsi get_negative_pair_col() pada SBERT_MNR.....	37
Gambar 3.24. Fungsi _generate_val_data() pada SBERT_MNR.....	37
Gambar 3.25. Fungsi _generate_train_data() pada SBERT_MNR.....	38
Gambar 3.26. Fungsi fine_tune() pada SBERT_MNR.....	39
Gambar 3.27. Fungsi inference() pada SBERT_MNR.....	40
Gambar 3.28. Fungsi __init__() pada LAM.....	41
Gambar 3.29. Fungsi _arr_to_df() pada LAM.....	42
Gambar 3.30. Fungsi fill_wrong_taxonomy() pada LAM.....	42
Gambar 3.31. Fungsi get_predictions() pada LAM.....	43
Gambar 3.32. Fungsi get_predictions_article() pada LAM.....	44
Gambar 3.33. Fungsi _export_prediction() pada LAM.....	45
Gambar 3.34. Fungsi predict() pada LAM.....	45
Gambar 3.35. Fungsi _get_evaluation_data() pada LAM.....	46
Gambar 3.36. Fungsi evaluate() pada LAM.....	46

Gambar 3.37. Fungsi <code>macro_evaluate()</code> pada LAM.....	47
Gambar 3.38. Diagram Proses Segmentasi dengan <i>human-in-the-loop</i>	50
Gambar 3.39. Proses Inisiasi Class serta Pemanggilan Fungsi Prediksi.....	51
Gambar 3.40. Excel Hasil Prediksi dengan Dua Sheet.....	51
Gambar 3.41. Antarmuka <i>Human-in-the-loop</i>	52
Gambar 3.42. Flowchart Antarmuka <i>Human-in-the-loop</i>	53
Gambar 3.43. Memasukan File Excel kedalam Antarmuka.....	54
Gambar 3.44. Tampilan Antarmuka setelah Data Diunggah.....	54
Gambar 3.45. Tampilan Antarmuka Jika Memilih Others.....	55
Gambar 3.46. Tampilan Antarmuka dengan Notifikasi Tersimpan.....	56
Gambar 3.47. Tampilan Antarmuka Setelah Menyimpan File Excel.....	57



DAFTAR LAMPIRAN

Lampiran A. Surat Pengantar MBKM - MBKM 01.....	68
Lampiran B. Kartu MBKM - MBM 02.....	69
Lampiran C. Daily Task MBKM - MBKM 03.....	70
Lampiran D. Lembar Verifikasi Laporan MBKM - MBKM 04.....	84
Lampiran E. Surat Penerimaan MBKM (LoA).....	85
Lampiran F. Lampiran Pengecekan Hasil Turnitin.....	86

