



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

RINGKASAN JUDUL YANG DIAJUKAN

*Semua kolom di bawah ini diisi oleh mahasiswa yang sudah mendapat judul

| | |
|---|--|
| Judul / Topik Skripsi | KLASIFIKASI PEKERJAAN DI INDONESIA YANG TERPENGARUH OLEH TEKNOLOGI <i>ARTIFICIAL INTELLIGENCE</i> BERDASARKAN <i>OCCUPATIONAL EXPOSURE SCORE</i> MENGGUNAKAN INDOBERT |
| Latar Belakang dan Penelitian Terdahulu | <p>Latar Belakang</p> <p>Dalam beberapa dekade terakhir, kecerdasan buatan (Artificial Intelligence) telah berkembang secara signifikan yang memiliki dampak positif pada berbagai sektor seperti keuangan (Maurya et al., 2024), kesehatan (Saxena et al., 2024), transportasi (Abduljabbar et al., 2019), dan industri industri (Peres et al., 2020). AI merupakan sistem yang dapat mengolah dan menganalisis data secara mandiri untuk mengambil keputusan dengan mengadopsi kecerdasan manusia. Dengan kemampuan analisis tersebut, kecerdasan buatan ini meringankan beban manusia karena banyak pekerjaan yang dapat dilakukan. Namun, kelebihan ini dapat memberikan efek sebaliknya yaitu berpotensi menggantikan manusia dalam dunia ketenagakerjaan (Zarifhonarvar, 2023).</p> <p>Berdasarkan data dari Kompas pada tahun 2023, penerapan AI diperkirakan akan mempengaruhi 17 sektor lapangan usaha di Indonesia. Sekitar 26.7 juta pekerja, atau 22.1% dari total tenaga kerja pada tahun 2021, dapat terbantu atau pekerjaannya lebih efisien dengan teknologi AI. Sektor informasi dan komunikasi diprediksi terkena dampak terbesar yaitu 58.1%, sedangkan sektor pertanian, kehutanan, dan perikanan terkena dampak terkecil sebesar 1.3%. Penggunaan AI dalam pekerjaan sehari-hari dapat mengurangi rata-rata jam kerja dari 8 jam menjadi 6 jam per hari, dengan 2 jam sisanya yang dapat ditangani oleh AI.</p> <p>Sementara itu, menurut pakar ekonomi Universitas Airlangga, Prof. Dr. Sri Herianingrum, perkembangan AI berpotensi memperbesar ketimpangan dan mengakibatkan hilangnya 40% pekerjaan, terutama pekerjaan dengan keterampilan rendah yang dapat diotomatisasi atau digantikan oleh teknologi, seperti di sektor jasa yang melibatkan tugas – tugas rutin dan berulang. Sebagai contoh, penggunaan AI dalam sektor perbankan, misalnya, dapat mengurangi kebutuhan tenaga kerja untuk tugas administratif dan layanan pelanggan. Oleh karena itu, Prof. Sri menekankan pentingnya perhatian terhadap dampak AI dan perlunya pelatihan serta pengembangan keterampilan bagi tenaga kerja untuk mempersiapkan mereka menghadapi perubahan dalam pasar kerja akibat kemajuan teknologi.</p> <p>Berbagai studi telah mengukur seberapa dalam AI menembus dunia kerja. Webb (2019) membandingkan teks paten AI dengan deskripsi pekerjaan dari ONET dan sensus AS (1960–2000), lalu menghitung berapa banyak kemampuan AI yang juga muncul sebagai tugas manusia. Hasilnya menunjukkan AI paling banyak membantu pekerjaan yang membutuhkan keterampilan tinggi dan berpotensi mengurangi kesenjangan upah. Felten et al. (2019) mengembangkan metrik AI Occupational Impact (AIOI) untuk mengukur sejauh mana pekerjaan terekspos oleh teknologi AI. Metrik ini dihitung dengan mengaitkan kemampuan AI, seperti pengenalan gambar dan terjemahan bahasa, ke keterampilan yang dibutuhkan dalam berbagai pekerjaan. Skor AIOI untuk setiap pekerjaan diperoleh dengan mengalikan skor kemampuan AI dengan seberapa umum dan penting keterampilan tersebut dalam pekerjaan tersebut, berdasarkan data O*NET. Hasilnya menunjukkan bahwa pekerjaan dengan skor AIOI tinggi cenderung mengalami sedikit peningkatan upah, tanpa perubahan signifikan dalam tingkat pekerjaan. Namun, temuan ini juga mengindikasikan bahwa AI berpotensi</p> |



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

memperlebar polarisasi pasar tenaga kerja, terutama antara pekerjaan dengan keterampilan tinggi dan rendah.

Tolan et al. (2021) memetakan 59 tugas umum ke 14 kemampuan kognitif, lalu mencocokkannya dengan 328 *benchmark* AI dengan menggunakan metode *bidirectional mapping* dan *indirect mapping*. Hasilnya potensi otomatisasi tertinggi pada tugas pemrosesan visual dan sensorimotor. Müller & Düsing (2023) menguji 12.000 lowongan bidang operasi bisnis, melatih BERT untuk mengekstrak tugas dan mengklasifikasikan kapan AI bisa menggantikan (0–5, 5–10, >10 tahun). Hasilnya model berhasil mengekstrak kalimat tugas dengan 88% tepat ekstraksi, dengan hasil klasifikasi sebesar 73% AI bisa mengerjakan tugas tersebut. Hasil prediksi menunjukkan 25% tugas siap digantikan AI dalam 5 tahun, 44% dalam 5–10 tahun, dan 31% lebih dari 10 tahun. Felten et al. (2023) fokus pada paparan pekerjaan, industri, dan wilayah terhadap model bahasa AI seperti ChatGPT. Hasilnya telemarketer dan guru paling terpapar, serta korelasi positif antara upah dan paparan AI. Zarifhonarvar (2023) meneliti risiko otomatisasi akibat ChatGPT pada klasifikasi ISCO, lalu memperkirakan 32,8% pekerjaan berisiko penuh, 36,5% parsial, dan 30,7% relatif aman.

Terakhir, Xu et al. (2025) menggunakan GPT-4 untuk memparafrase 13.199 tugas dari ONET, ESCO, dan data pekerjaan Australia, lalu melatih BERT + softmax untuk memprediksi otomatisasi tiap tugas—hasilnya 25,1% tugas O*NET berisiko tinggi, terutama di perhotelan & kuliner, administrasi & pendukung, ritel, pertambangan, dan manufaktur. Berdasarkan penelitian-penelitian sebelumnya, pendekatan yang digunakan dalam mengukur dampak AI terhadap pekerjaan bervariasi, mulai dari metode berbasis aturan (rule-based) hingga pemanfaatan model BERT. Namun, sebagian besar studi tersebut belum mempertimbangkan konteks lokal Indonesia.

Seperti yang kita ketahui, penggunaan AI di industri terus meningkat di Indonesia, namun dampaknya terhadap sektor pekerjaan lokal di Indonesia belum terdokumentasi secara mendalam. Sebagian besar penelitian yang ada mengenai dampak teknologi AI terhadap pekerjaan berasal dari negara Barat dan China yang tentunya tidak berlaku seluruhnya terhadap pasar kerja Indonesia yang unik seperti faktor literasi digital dan struktur ekonomi Indonesia. Oleh karena itu, dibutuhkan penelitian dalam sektor lokal untuk memahami dampak teknologi AI terhadap pekerjaan di Indonesia

IndoBERT adalah model bahasa yang dirancang khusus untuk menangani korpus bahasa Indonesia. IndoBERT unggul dalam menangkap nuansa leksikal, sintaksis, dan terminologi khas yang sering luput dari model multilingual atau pendekatan rule-based. Koto et al. (2020) memperkenalkannya sebagai pre-trained model pertama untuk NLP Bahasa Indonesia, menunjukkan embedding yang lebih kaya konteks; selanjutnya, fine-tuning IndoBERT pada domain baru dengan menyesuaikan bobot internal dan lapisan klasifikasi tambahan yang dapat meningkatkan akurasi tugas klasifikasi hingga 98–99% (Baharuddin & Naufal, 2023; Naufal, 2022), termasuk klasifikasi soal ujian lokal dan deskripsi pekerjaan yang terpengaruh AI. Hal ini menjadikan IndoBERT lebih andal dalam memahami beragam bahasa Indonesia mulai dari struktur formal dalam dokumen hingga kosakata dan gaya informal di teks lowongan kerja.

Di Indonesia, adopsi teknologi Kecerdasan Buatan (AI) telah mengalami perkembangan pesat dalam beberapa tahun terakhir. Salah satu penerapan AI yang menonjol adalah penggunaan dalam penegakan hukum lalu lintas elektronik di Yogyakarta (Rahmat et.al., 2021). Kecerdasan buatan memungkinkan mesin dapat melakukan pekerjaan yang biasanya dikerjakan oleh manusia, sehingga dapat meningkatkan efisiensi dan produktivitas. Namun, di sisi lain, kehadiran AI juga memicu kekhawatiran akan dampak negatifnya terhadap lapangan pekerjaan. Salah satu aspek yang menjadi perhatian adalah potensi pengurangan tenaga kerja



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

manusia di berbagai bidang. Saat ini, AI telah mampu melakukan berbagai tugas, mulai dari pengolahan gambar, penulisan konten, hingga pemecahan masalah matematis dan filosofis. Dengan kemampuan yang semakin canggih, AI dapat menggantikan pekerjaan-pekerjaan yang sebelumnya dilakukan oleh manusia. Dampak ini dapat terasa di berbagai sektor, seperti manufaktur, layanan, dan bahkan pekerjaan-pekerjaan intelektual. Hal ini dapat mengakibatkan pengangguran dan ketimpangan ekonomi yang semakin besar di masyarakat.

Penelitian ini berfokus pada klasifikasi pekerjaan di Indonesia yang terpengaruh oleh teknologi AI dengan menerapkan Natural Language Processing (NLP). Proses ini mencakup tahapan text preprocessing pada 3 dataset berbasis teks yaitu dataset kemampuan AI (dari Google Patents), dan dataset pekerjaan (dari ISCO-08 dan LinkedIn Jobs yang berlokasi di Indonesia). Setelah itu, dilanjutkan dengan perhitungan Occupational Exposure Score yaitu jumlah verb noun pairs yang tumpang tindih (overlap) antara kemampuan AI dan kemampuan yang dibutuhkan oleh pekerjaan manusia. Setiap gabungan kata yang overlap akan disimpan dan dihitung jumlahnya. Maka, nilai eksposur dapat ditentukan melalui jumlah kata yang overlap tersebut. Berdasarkan skor tersebut, pekerjaan akan dikategorikan menjadi 3 label, yaitu Terotomasi penuh (digantikan AI) dengan skor 5-7 overlap, Terotomasi sebagian (dibantu AI) dengan skor 1-4 overlap, dan Tidak terpengaruh AI dengan skor 0 overlap.

Penelitian ini akan mengembangkan dashboard interaktif yang dirancang untuk mengidentifikasi dan mengklasifikasikan tingkat pengaruh suatu pekerjaan oleh teknologi AI. Fitur utama dari dashboard ini adalah menu "Jobs Testing", yang memungkinkan pengguna untuk menginput nama dan deskripsi pekerjaan tertentu. Sistem kemudian akan memproses input tersebut melalui tahapan text preprocessing, menghitung jumlah kata overlap, dan mengklasifikasikan pekerjaan tersebut apakah masuk ke dalam tiga kategori: Terotomasi Penuh, Terotomasi Sebagian, atau Tidak Terpengaruh AI dengan hasil output akhir yang mencakup kolom seperti Nama Pekerjaan, Deskripsi Pekerjaan, Kata Overlap, dan Hasil Klasifikasi.

Selain fitur pengujian pekerjaan, dashboard ini juga direncanakan dapat menyajikan gambaran umum data serta visualisasi statistik klasifikasi pekerjaan di berbagai sektor di Indonesia. Tujuannya adalah untuk memberikan dasar untuk merancang solusi strategis dengan menyusun rekomendasi kebijakan guna meningkatkan kompetensi sumber daya manusia dan mendorong inovasi dalam transformasi pekerjaan di Indonesia yang terdampak oleh AI.

Penelitian Terdahulu

| No. | Penulis | Judul | Tahun |
|-----|---|--|-------|
| 1. | Michael Webb | The Impact of Artificial Intelligence on the Labor Market | 2019 |
| 2. | Edward W. Felten, Manav Raj, Robert Seamans | The Occupational Impact of Artificial Intelligence : Labor, Skills, and Polarization | 2019 |



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

| | | | |
|----|--|--|------|
| 3. | Songül Tolan, Annarosa Pesole, Fernando Martínez-Plumed, Enrique Fernández-Macías, José Hernández-Orallo, Emilia Gómez | Measuring the Occupational Impact of AI: Tasks, Cognitive Abilities and AI Benchmarks | 2021 |
| 4. | Ali Zarifhonarvar | Economics of ChatGPT: a labor market view on the occupational impact of artificial intelligence | 2023 |
| 5. | Edward W. Felten, Manav Raj, Robert Seamans | How will language modelers like ChatGPT affect occupations and industries | 2023 |
| 6. | Oliver Müller, Christoph Düsing | Forecasting the Impact of Artificial Intelligence on the Labor Market Through Natural Language Processing of Job Posting | 2023 |
| 7. | Xu et al. | From Occupations to Tasks: A New Perspective on Automatability Prediction Using BERT | 2025 |

Perbedaan Penelitian :

Penelitian yang akan dilakukan banyak mengadopsi penelitian (Webb, 2019) dengan beberapa modifikasi. Pertama, mempersempit wilayah menjadi skala lokal yakni hanya pekerjaan di wilayah Indonesia saja. Sumber data yang digunakan untuk mengamati kemampuan AI yaitu data dari Google Patent karena data ini terus diperbaharui sehingga tetap relevan dengan kemampuan AI saat ini. Sementara itu, data mengenai kebutuhan kemampuan pekerjaan diperoleh dari standarisasi klarifikasi pekerjaan terbaru yaitu ISCO 08 dan LinkedIn Jobs yang berlokasi di Indonesia (dikarenakan dataset pada ISCO-08 tidak memuat pekerjaan baru yang lagi *booming* saat ini seperti data analyst, data scientist, dan sebagai nya).

Pada penelitian Webb sebelumnya hanya berbasis metode NLP *Rule-Based*, tetapi pada penelitian ini menggunakan IndoBERT sebagai model untuk melakukan klasifikasi pekerjaan yang terpengaruh oleh teknologi AI.



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI

UNIVERSITAS SUMATERA UTARA FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

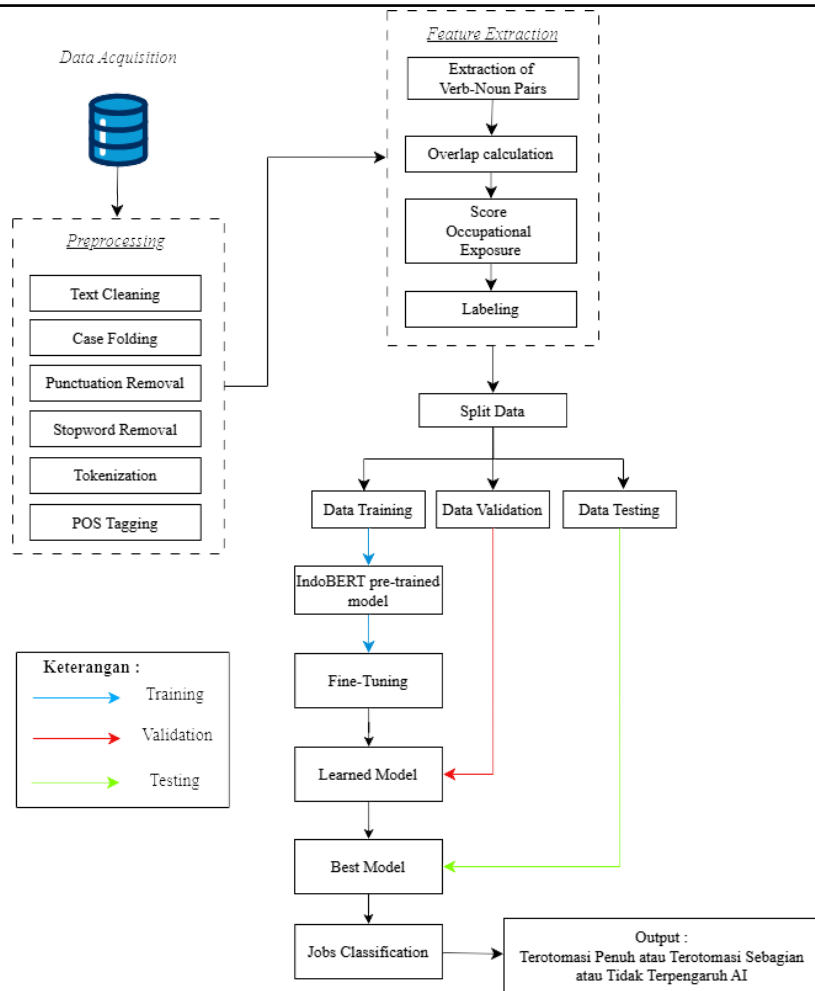
PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

Rumusan Masalah

Perkembangan pesat kecerdasan buatan (AI) telah mendorong perubahan besar di dunia kerja sehingga menimbulkan kekhawatiran soal stabilitas dan kualitas lapangan pekerjaan di Indonesia. teknologi ini juga mulai menggantikan tugas-tugas rutin yang sebelumnya dikerjakan manusia seperti kasir di toko ritel dengan sistem self-checkout, penerjemah bahasa otomatis, pengemudi taksi lewat mobil otonom, serta customer service yang beralih ke chatbot; bahkan content writer dasa. Di sisi lain, pasar kerja Indonesia masih menunjukkan ketidakstabilan dengan tingginya angka setengah pengangguran dan pekerja paruh waktu, yang mengindikasikan adanya tekanan struktural akibat transformasi digital. Namun, sistem klasifikasi pekerjaan yang digunakan saat ini (KBJI 4-digit) belum mampu memetakan tingkat kerentanan terhadap otomatisasi, sehingga pemerintah dan pemangku kepentingan kesulitan merancang langkah yang tepat. Hal ini menjadi pemicu dibutuhkan suatu pendekatan yang efektif untuk mengklasifikasikan pekerjaan berdasarkan tingkat paparan terhadap teknologi AI agar mendukung pengambilan keputusan yang lebih akurat dalam pengembangan sumber daya manusia.

Metodologi



Arsitektur umum dalam penelitian ini dapat dilihat pada gambar diatas yang dirancang untuk memberi gambaran proses klasifikasi pekerjaan yang terpengaruh oleh teknologi AI menggunakan model IndoBERT. Proses ini terdiri dari beberapa tahapan mulai dari data *acquisition*, data *preprocessing*, *feature extraction*, model development, dan output klasifikasi. Berikut adalah penjelasan rinci dari setiap tahapan:



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

1. Data Acquisition

Data acquisition atau pengumpulan data berupa data patent yang berisi kemampuan AI dan data pekerjaan manusia yang berlokasi di Indonesia yang akan dijadikan dataset dalam pengembangan model. Adapun dataset yang digunakan adalah sebagai berikut.

a. Google Public Patents Data

Google Public Patent Data merupakan dataset *open-source* yang dapat diakses pada Google Cloud melalui link berikut :

https://console.cloud.google.com/marketplace/product/google_patents_public_datasets/google-patents-public-data?project=tidal-theater-380309

Data yang diambil yaitu data terkait *artificial intelligence* menggunakan *query* SQL dengan *keyword* “*machine learning*”, “*computer vision*”, “*robotic*”, “*natural language processing*”, “*automated planning and scheduling*”, “*deep learning*”, “*neural network*”, dan “*artificial intelligence*” pada atribut *title_localized* dan *abstract_localized* dengan pertimbangan bahwa *title_localized* berisi penamaan untuk kemampuan AI, sedangkan *abstract_localized* berisi deskripsi semua kemampuan yang dimiliki AI dalam tugas tersebut. Oleh karena itu, kedua fitur ini diperlukan untuk mengekstrak kemampuan dan tugas yang dapat dikerjakan AI.

Processing location: US X

Query results

| Job information | | Results | Chart | JSON | Execution details | Execution graph |
|-----------------|------------------|--------------|---------------------------------|--|-------------------|-----------------|
| Row | publication_date | country_code | title | abstract | | |
| 1 | 19870602 | US | Artificial intelligence system | An artificial intelligence system for accepting a statement, understanding the statement and making a response to the | | |
| 2 | 19880531 | US | Computer vision system based... | A system for computer vision in association with a computer comprises a solid state image sensor or optic ram, a lens, a clock | | |
| 3 | 19880621 | US | Temporal sequences with neur... | A sequence generator employing a neural network having its output coupled to at least one plurality of delay elements. The | | |
| 4 | 19880726 | US | Neural networks | Neural network type information processing | | |

Alasan pemilihan keyword tersebut karena *deep learning*, *machine learning*, dan *neural network* menjadi algoritma utama yang digunakan dalam teknologi AI (Choi et al., 2020), sehingga data terkait *keyword* tersebut akan secara langsung mengarah pada AI. Selain itu, *computer vision* dan *natural language processing* merupakan cabang AI dimana *computer vision* menganalisis data visual baik gambar maupun video, sedangkan NLP menganalisis data teks dan bahasa manusia (Guo et al., 2020).

Setelah melakukan penyaringan berdasarkan *keyword* di atas, dilakukan juga *filter* untuk mengambil *keyword* row data yang berbahasa Inggris (EN) dan berlokasi di US untuk menghindari bahasa asing yang tidak relevan.



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

b. ISCO-08

Dataset ISCO-08 (International Standard Classification of Occupations) dikembangkan oleh badan International Labour Organization (ILO) yang menyediakan klasifikasi pekerjaan di seluruh dunia beserta deskripsinya. ISCO-08 dapat diakses melalui link <https://isco-ilo.netlify.app/en/isco-08/>. Atribut yang digunakan yaitu *Title EN* yang berisi nama profesi dan *Tasks include* yang berisi tugas dan kemampuan yang harus dimiliki seseorang dalam profesi tersebut.

c. LinkedIn Jobs

Data dari platform LinkedIn akan dikumpulkan menggunakan teknik *web scraping*, yaitu metode otomatis untuk mengekstrak informasi dari situs web. Penulis memfilter lokasi “Indonesia”, lalu mengambil judul dan deskripsi pekerjaan, dan menyimpannya dalam format CSV.

Dataset dari ISCO-08 kemudian akan digabung dengan hasil scraping LinkedIn berdasarkan kolom Nama Pekerjaan dan Deskripsi Pekerjaan. Seluruh data, termasuk kemampuan AI dan pekerjaan, akan diterjemahkan ke dalam bahasa Indonesia agar lebih relevan dengan pasar kerja Indonesia dan meningkatkan akurasi analisis.

2. Data Preprocessing

Setelah data dikumpulkan, tahap pertama adalah *text cleaning*, yaitu menghapus *missing value* dan substring yang tidak relevan. Tahap kedua adalah *case folding*, yakni menyamakan seluruh huruf menjadi huruf kecil agar perhitungan kata tidak terpengaruh perbedaan kapitalisasi.

Tahap ketiga, *punctuation removal*, dilakukan untuk menghapus tanda baca agar mengurangi noise dan fokus pada kata bermakna. Tahap keempat adalah *stopword removal*, yaitu menghapus kata-kata umum seperti “dan”, “yang”, dan “di” yang tidak memberikan makna signifikan dalam analisis.

Tahap kelima, *tokenization*, membagi teks menjadi unit kata atau frasa, misalnya “menjaga keamanan” menjadi “menjaga” dan “keamanan”, untuk memudahkan identifikasi dan pelabelan.

Tahap terakhir adalah POS (*Part-of-Speech*) *tagging*, yaitu memberi label jenis kata (kata benda, kerja, sifat, atau keterangan) berdasarkan konteks kalimat (Chiche, 2022), guna memahami struktur dan makna kata dalam teks.

3. Feature Extraction

Setelah data dibersihkan, tahapan ketiga adalah *feature extraction*, dimulai dengan *extraction of verb-noun pairs*. Tahapan ini bertujuan mengambil kata kunci paling relevan yang merepresentasikan aksi (verb) yang dapat dilakukan AI terhadap objek (noun). Ekstraksi dilakukan menggunakan algoritma *dependency parsing*, yang mengidentifikasi hubungan sintaksis antar kata dalam kalimat (Honnibal & Johnson, 2015), dengan akurasi 91,85% berdasarkan tolak ukur standar NLP.

Tahapan ini dilakukan setelah *POS tagging*, di mana setiap kata diberi label gramatikal. Pola *dependency parsing* yang digunakan mengelompokkan kata kerja dan objek langsungnya menjadi pasangan *verb-noun*. Parsing membentuk struktur



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

pohon yang menggambarkan hubungan seperti subjek (nsubj) dan objek langsung (dobj) terhadap kata kerja utama (ROOT).

Contoh:

Kalimat: “Dokter merawat pasien”

POS Tagging: “Dokter” = NOUN, “Merawat” = VERB, “Pasien” = NOUN

Dependency: “merawat” = ROOT, “dokter” = nsubj, “pasien” = dobj

Hasil ekstraksi: (merawat, pasien)

Pendekatan ini serupa dengan penelitian Webb (2019), yang mengekstrak *verb-noun pairs* dari deskripsi pekerjaan dan paten AI untuk mengukur kesamaan tugas. Tahapan kedua dari *feature extraction* adalah *overlap calculation*, yakni menghitung jumlah pasangan *verb-noun* yang muncul di kedua sumber: kemampuan AI dan deskripsi pekerjaan manusia. Nilai *occupational exposure* ditentukan dari jumlah pasangan yang *overlap*.

Contoh:

Kemampuan AI: “mendeteksi kerusakan”, “mengidentifikasi objek”, “menganalisis data”

Deskripsi pekerjaan: “mendeteksi kerusakan”, “mengevaluasi kualitas”, “menganalisis data”

Overlap: 2 pasangan → “mendeteksi kerusakan”, “menganalisis data”

Semakin tinggi nilai *overlap*, semakin besar potensi pekerjaan terdampak oleh AI. Menurut Webb (2019), dampak diklasifikasikan menjadi:

- Tidak terdampak (0 overlap)
- Terdampak sebagian (1–4 overlap)
- Terdampak penuh (5–7 overlap)

4. Model Development

Pada tahapan ini, dimulai dari *split data* dimana dataset dibagi menjadi data *training* (70%), data *validation* (10%), dan data *testing* (20%) yang kemudian akan ditransformasi setiap teks menjadi vektor menggunakan IndoBERT *pre-trained* model, yaitu model bahasa yang telah dilatih sebelumnya khusus untuk bahasa Indonesia. Dengan memanfaatkan pengetahuan awal tentang struktur dan kosakata bahasa Indonesia, IndoBERT memberikan konteks yang kaya pada setiap token.

Setelah vektor *embedding* dihasilkan, dilakukan proses *fine-tuning*. Menurut (Naufal, 2022) dalam *Fine-Tuning IndoBERT*, parameter IndoBERT yang semula umum akan dioptimasi lebih lanjut agar model dapat mengklasifikasikan pekerjaan secara akurat. Tahap ini melibatkan penyesuaian bobot internal model dan penambahan lapisan klasifikasi. Dengan demikian, model hasil *fine-tuning* akan lebih tepat untuk memprediksi label terotomasi penuh, terotomasi sebagian, atau tidak terpengaruh AI. Data *validation* akan digunakan pada *learned model* dan untuk memantau metrik seperti loss dan akurasi serta mencegah terjadinya overfitting. Hasil akhirnya dievaluasi menggunakan data *testing* untuk memastikan performa akhir model.

5. Output Klasifikasi

Model yang sudah jadi akan diterapkan pada dashboard untuk implementasi. Pengguna dapat mengunggah nama pekerjaan dan deskripsi pekerjaan dalam teks berbahasa Indonesia. Sistem kemudian akan memproses input tersebut melalui tahapan *text preprocessing*, menghitung jumlah *kata overlap*, dan



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

| | |
|-----------|---|
| | mengklasifikasikan pekerjaan tersebut apakah masuk ke dalam tiga kategori: Terotomasi Penuh, Terotomasi Sebagian, atau Tidak Terpengaruh AI dengan hasil output akhir yang mencakup kolom seperti Nama Pekerjaan, Deskripsi Pekerjaan, Kata Overlap, dan Hasil Klasifikasi. |
| Referensi | <p>Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. A. (2019). Applications of artificial intelligence in transport: An overview. <i>Sustainability</i>, 11(1), 189. https://doi.org/10.3390/su11010189</p> <p>Baharuddin, F., & Naufal, M. F. (2023). Fine-tuning IndoBERT for Indonesian exam question classification based on bloom's taxonomy. <i>Journal of Information Systems Engineering and Business Intelligence</i>, 9(2), 253-263. https://doi.org/10.20473/jisebi.9.2.253-263</p> <p>Chiche, A., & Yitagesu, B. (2022). Part of speech tagging: A systematic review of deep learning and machine learning approaches. <i>Journal of Big Data</i>, 9(1). https://doi.org/10.1186/s40537-022-00561-y</p> <p>Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to Machine Learning, Neural Networks, and Deep Learning. <i>Translational vision science & technology</i>, 9(2), 14. https://doi.org/10.1167/tvst.9.2.14</p> <p>Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In <i>Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)</i> (pp. 4171-4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423.</p> <p>Felten, E. W., Raj, M., & Seamans, R. (2019). The variable impact of artificial intelligence on labor: The role of complementary skills and technologies. <i>SSRN Electronic Journal</i>. https://doi.org/10.2139/ssrn.3368605</p> <p>Felten, E. W., Raj, M., & Seamans, R. (2023). How will language modelers like ChatGPT affect occupations and industries? <i>SSRN Electronic Journal</i>. https://doi.org/10.2139/ssrn.4375268</p> <p>Guo, J., He, H., He, T., Lausen, L., Li, M., Lin, H., Shi, X., Wang, C., Xie, J., Zha, S., Zhang, A., Zhang, H., Zhang, Z., Zhang, Z., Zheng, S., & Zhu, Y. (2020). GluonCV and gluon NLP: Deep learning in computer vision and natural language processing. <i>Journal of Machine Learning Research</i>, 21(1), 845-851.</p> |



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

- Honnibal, M., & Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. <https://doi.org/10.18653/v1/d15-1162>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. *Proceedings of the 28th International Conference on Computational Linguistics*. <https://doi.org/10.18653/v1/2020.coling-main.66>
- Maurya, S., Verma, R., Khilnani, L., Bhakuni, A. S., Kumar, M., & Rakesh, N. (2024). Effect of AI on the financial sector: Risk control, investment decision-making, and business outcome. *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 1-7. <https://doi.org/10.1109/icrito61523.2024.10522410>
- Müller, Oliver and Düsing, Christoph, Forecasting the Impact of Artificial Intelligence on the Labor Market Through Natural Language Processing of Job Posting (June 19, 2023). <http://dx.doi.org/10.2139/ssrn.4710299>
- Peres, R. S., Jia, X., Lee, J., Sun, K., Colombo, A. W., & Barata, J. (2020). Industrial artificial intelligence in industry 4.0 - Systematic review, challenges and outlook. *IEEE Access*, 8, 220121-220139. <https://doi.org/10.1109/access.2020.3042874>
- Saxena, A. K., Ness, S., & Khinvasara, T. (2024). The influence of AI: The revolutionary effects of artificial intelligence in healthcare sector. *Journal of Engineering Research and Reports*, 26(3), 49-62. <https://doi.org/10.9734/jerr/2024/v26i31092>
- Sharma, A., & Parmar, M. (2021). A survey on text pre-processing and feature extraction techniques for sentiment analysis of Twitter data. *International Research Journal of Computer Science*, 8(12), 271-278. <https://doi.org/10.26562/irjcs.2021.v08i2.001>
- Rahmat, A. F., & Pribadi, U. (2021). Delivering artificial intelligence for electronic traffic law enforcement in Yogyakarta region: Current effort and future challenges. *IOP Conference Series: Earth and Environmental Science*, 717(1), 012016. <https://doi.org/10.1088/1755-1315/717/1/012016>
- Rosalina, M. P., Krisna, A., & Wisanggeni, S. P. (2023, June 28). *Dampak AI terhadap Lapangan Usaha Di Indonesia*. [kompas.id](https://www.kompas.id/baca/investigasi/2023/06/27/dampak-ai-di-lapangan-usaha). <https://www.kompas.id/baca/investigasi/2023/06/27/dampak-ai-di-lapangan-usaha>
- Tabassum, A., & Patil, D.R. (2020). A Survey on Text Pre-Processing & Feature



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

Extraction Techniques in Natural Language Processing. *International Research Journal of Engineering and Technology (IRJET)*, 7(06), 4864-4867.

Tolan, S., Pesole, A., Martínez-Plumed, F., Fernández-Macías, E., Hernández-Orallo, J., & Gómez, E. (2021). Measuring the occupational impact of AI: Tasks, cognitive abilities and AI benchmarks. *Journal of Artificial Intelligence Research*, 71, 191-236. <https://doi.org/10.1613/jair.1.12647>

Unairnews. (2024, April 19). *AI Mengancam Stabilitas Ketenagakerjaan? Begini Kata Pakar Ekonomi*. <https://unair.ac.id/ai-mengancam-stabilitas-ketenagakerjaan-begini-kata-pakar-unair/>

Webb, M. (2019). The impact of artificial intelligence on the labor market. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3482150>

Yogish, D., Manjunath, T. N., & Hegadi, R. S. (2019). Review on natural language processing trends and techniques using NLTK. In *Recent Trends in Image Processing and Pattern Recognition: Second International Conference, RTIP2R 2018, Solapur, India, December 21–22, 2018, Revised Selected Papers, Part III-2* (pp. 589-606). Springer Singapore.

Zarifhonarvar, A. (2024), "Economics of ChatGPT: a labor market view on the occupational impact of artificial intelligence", *Journal of Electronic Business & Digital Economics*, Vol. 3 No. 2, pp. 100-116. <https://doi.org/10.1108/JEBDE-10-2023-0021>.

Medan, 09 Mei 2025
Mahasiswa yang mengajukan,

Kezia Natalia
NIM. 211402002