

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Sebagai salah satu Perguruan Tinggi Vokasi, Politeknik Negeri Jember (POLIJE) memiliki kewajiban untuk memastikan bahwa lulusan yang dihasilkan memiliki keahlian yang tinggi dan relevansi. Program studi Manajemen Informatika (MIF) memiliki tiga profil lulusan utama, yakni *Programmer*, *Data Analyst*, dan Wirausaha Informatika yang secara berkala digunakan untuk memetakan penyerapan lulusan secara objektif sesuai kompetensi inti program studi. Salah satu instrumen yang digunakan dalam evaluasi ini adalah *tracer study*, yang menurut Cruz (2022) dapat digunakan untuk mendefinisikan lebih lanjut misi, ceruk pasar sebuah Institusi Pendidikan Tinggi, dan menyoroti bagaimana program akademik serta penawaran mata kuliah dapat diubah untuk menyesuaikan dengan tujuan institusional. Namun, penggunaan *tracer study* saat ini belum ideal, terutama dalam hal mengklasifikasikan pekerjaan alumni berdasarkan profil lulusan yang telah ditetapkan.

Analisis lebih lanjut menemukan bahwa sumber masalah adalah proses pengolahan data *tracer study* masih dilakukan secara manual, terutama dalam pengkategorian dan pemetaan pekerjaan alumni ke profil lulusan. Proses yang dilakukan secara manual ini rentan akan inkonsistensi data dari tahun ke tahun dan menyulitkan audit secara internal maupun eksternal. Hal ini mengakibatkan kesenjangan kompetensi antara lulusan dengan kebutuhan industri. Terlebih lagi, kualitas data yang tidak akurat berpotensi menurunkan akreditasi, mengingat *tracer study* merupakan syarat penting untuk akreditasi perguruan tinggi (Romadlon & Arifin, 2021). Secara strategis, data yang tidak akurat dan proses yang tidak standar ini sangat berbahaya, sebab menjadikannya tidak relevan untuk evaluasi dan revisi kurikulum. Kondisi ini menghambat kemampuan program studi untuk melakukan pengambilan keputusan berbasis data, yang sangat penting untuk menjaga relevansi dengan persyaratan industri. Pada akhirnya, ketidakakuratan data penyerapan lulusan dapat mengurangi kepercayaan industri terhadap kompetensi lulusan program studi.

Untuk mengatasi kelemahan fundamental dari proses manual tersebut, penelitian ini mengusulkan pengembangan sistem informasi yang mampu melakukan klasifikasi pekerjaan alumni secara otomatis dan objektif. Sistem ini dirancang untuk menerima data *tracer study* dalam bentuk file *Excel* yang bersumber dari *tracer study* internal MIF yang diolah lebih terstruktur dan siap diproses secara *batch*. Selain mengolah data internal, sistem ini juga dirancang untuk mengintegrasikan data Tracer Study resmi dari Kementerian Pendidikan Tinggi, Sains, dan Teknologi (Kemendiktisaintek). Langkah integrasi ini diperlukan karena data Kemendiktisaintek memuat parameter standar nasional yang kaya akan informasi pelaporan akreditasi, seperti detail jenis instansi, skala pendapatan, dan tingkat keselarasan. Kombinasi kedua sumber data ini akan memberikan potret penyerapan lulusan yang menyeluruh pada dasbor visual.

Namun, mengingat volume data internal yang tersedia relatif kecil (kurang dari 400 sampel berlabel), sistem ini dibangun menggunakan arsitektur hybrid. Sistem menerapkan *rule-based keyword matching* sebagai klasifikasi tahap pertama dan *Multinomial Logistic Regression* dengan ekstraksi fitur TF-IDF sebagai fallback untuk teks ambigu, ditambah mekanisme *confidence threshold* untuk mengarahkan prediksi berkeyakinan rendah ke antrean *review manual*. Pilihan menggunakan *Multinomial Logistic Regression* didukung oleh temuan Ortiz dkk. (2022) yang secara spesifik membuktikan bahwa algoritma klasik berbasis *Bag-of-Words*/TF-IDF mampu mengungguli arsitektur deep learning pada tugas klasifikasi teks pekerjaan, terutama ketika *dataset* bersifat terbatas dan tidak seimbang. Selain itu, studi komparatif Hamdi dkk. (2026) pada pemrosesan teks berbahasa Indonesia, meskipun dalam domain analisis sentimen memvalidasi bahwa model klasik seperti *Logistic Regression* menawarkan stabilitas prediksi dan efisiensi komputasi yang signifikan dibanding arsitektur kompleks seperti BiLSTM. Kesamaan karakteristik linguistik antara kedua konteks tersebut terletak pada sifat teks yang singkat, mengandung variasi diksi tidak baku, dan memerlukan representasi fitur yang ringan. Oleh karena itu, pendekatan klasik dipilih bukan karena kesamaan domain tugas, melainkan karena efisiensi inferensi, kemudahan interpretasi probabilitas,

dan ketahanan terhadap noise morfologis bahasa Indonesia yang konsisten terbukti pada korpus teks singkat.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

- a. Bagaimana cara mengembangkan sistem informasi *tracer study* berbasis web yang menerapkan metode TF-IDF untuk ekstraksi fitur teks dari data pekerjaan dan menggunakan algoritma *Multinomial Logistic Regression* untuk mengklasifikasikannya sesuai dengan profil lulusan Program Studi MIF?
- b. Bagaimana merancang dan membangun visualisasi data yang interaktif dan informatif dalam sistem *tracer study* berbasis web untuk mendukung pengambilan keputusan dalam menganalisis penyerapan alumni lulusan Program Studi MIF?
- c. Bagaimana mengevaluasi performa sistem klasifikasi pada data internal MIF dan mengimplementasikan mekanisme mitigasi (*confidence threshold* dan *manual review*) untuk menjaga akuntabilitas hasil klasifikasi?

1.3 Batasan Masalah

Dari rumusan masalah yang ada di atas, dapat diambil batasan masalah sebagai berikut:

- a. Sistem ini hanya melibatkan dua jenis pengguna, yakni Ketua Program Studi Manajemen Informatika dan pihak CDC (*Career Development Center*).
- b. Proses klasifikasi dibatasi pada data pekerjaan alumni hasil dari *tracer study* Program Studi Manajemen Informatika Politeknik Negeri Jember.
- c. Perolehan data untuk proses klasifikasi dibatasi dari hasil *tracer study* Program Studi Manajemen Informatika Politeknik Negeri Jember dari tahun 2021 sampai 2023.
- d. Sistem menerapkan pendekatan *hybrid* (*rule-based* + *Multinomial Logistic Regression*) dengan *confidence threshold* 0.50. Prediksi di bawah *threshold* akan

dialihkan ke (*queue*) tinjauan (*review*) manual oleh admin/CDC untuk memastikan akurasi pelaporan akreditasi.

- e. Integrasi data eksternal dibatasi pada ekstraksi dan visualisasi data dari format file *Tracer Study* Kemendiktisaintek untuk memperkaya visualisasi dasbor, tanpa melalui proses retraining model *Machine Learning* internal.

1.4 Tujuan

Dengan diuraikannya latar belakang masalah tersebut, maka tujuan dari tugas akhir ini adalah:

- a. Mengembangkan sistem informasi *tracer study* berbasis web yang menerapkan metode TF-IDF untuk ekstraksi fitur teks dari data pekerjaan dan menggunakan algoritma *Multinomial Logistic Regression* untuk mengklasifikasikannya sesuai dengan profil lulusan Program Studi MIF.
- b. Merancang dan membangun visualisasi data yang interaktif dan informatif dalam sistem *tracer study* berbasis web untuk mendukung pengambilan keputusan dalam menganalisis penyerapan alumni lulusan Program Studi MIF.
- c. Mengevaluasi performa sistem klasifikasi secara eksklusif pada data internal MIF dan mengimplementasikan mekanisme mitigasi berbasis *confidence threshold* serta antrean *review manual*.

1.5 Manfaat

Dengan tujuan dibuatnya tugas akhir ini, diharapkan memberikan manfaat sebagai berikut:

- a. Memberikan kontribusi berupa studi kasus implementasi sistem klasifikasi teks dari hulu ke hilir, mulai dari pra-pemrosesan data, ekstraksi fitur menggunakan TF-IDF, hingga penerapan algoritma *Multinomial Logistic Regression* pada data pekerjaan alumni. Hasil yang didapatkan dapat menjadi dasar pengembangan metode klasifikasi otomatis berbasis teks dalam konteks pendidikan tinggi dan manajemen data alumni.
- b. Mampu mengotomatisasi proses klasifikasi pekerjaan alumni berdasarkan data teks yang direkap dari data *Tracer Study* Internal MIF, serta mengintegrasikan

data *Tracer Study* Kemendiktisaintek ke dalam modul agregasi dasbor. Dengan adanya proses otomatisasi ini, pihak program studi tidak perlu mengkategorikan data pekerjaan alumni secara manual, yang selama ini beresiko menghasilkan data yang tidak akurat. Sistem ini juga dilengkapi dengan visualisasi data yang interaktif dan informatif dengan tujuan mempermudah proses analisis dan pengambilan keputusan.

- c. Menyediakan mekanisme *fallback* berbasis *confidence threshold* dan anotasi manual yang memungkinkan CDC mengoreksi prediksi ambigu, sehingga sistem tetap menghasilkan laporan akreditasi yang valid meskipun menghadapi keterbatasan data pelatihan.
- d. Sistem ini dapat membantu Program Studi MIF POLIJE dalam:
 - 1) Memperoleh data alumni yang lebih terstruktur, konsisten, dan mudah dianalisis.
 - 2) Mengevaluasi kesesuaian antara profil lulusan dan pekerjaan alumni secara lebih mendalam.
 - 3) Mendukung proses akreditasi program studi dengan menyediakan data yang akurat.