

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Teknologi berkembang pesat dalam dekade ini, dengan banyak sektor, seperti jual beli, beralih dari tradisional menjadi dalam jaringan (daring). Jurusan Informatika di Indonesia semakin populer seiring dengan perkembangan tersebut. Hal ini menunjukkan adanya potensi besar di bidang teknologi yang dapat dimanfaatkan di masa depan.

Menurut data yang disajikan oleh Sekretariat Direktorat Jenderal Pendidikan Tinggi pada tahun 2022, jurusan teknik informatika menempati posisi empat teratas sebagai jurusan program studi lulusan terbanyak [1]. Tidak heran jika jurusan ini menarik begitu banyak peminat, dunia digital yang terus berevolusi menawarkan peluang karir yang luas dan menjanjikan bagi lulusan Teknik Informatika.

Kebutuhan tenaga kerja di sektor TIK diproyeksikan terus meningkat, dari 1.232.666 pada 2022 menjadi 1.979.418 pada 2025. Meskipun minat terhadap jurusan teknologi tinggi, jumlah lulusan masih belum mencukupi proyeksi kebutuhan tenaga kerja di sektor ini [2].

Rekognisi pembelajaran lampau memungkinkan pengalaman belajar non-formal, seperti pelatihan atau sertifikasi, diakui sebagai kompetensi relevan. Penelitian ini mengklasifikasikan sertifikat berdasarkan tujuh mata kuliah jurusan Informatika untuk mempermudah konversi keahlian non-formal ke dalam program akademik.

Metode machine learning (ML) tradisional seperti SVM, KNN, dan K-Means sering digunakan untuk klasifikasi teks. Namun, untuk hasil yang lebih akurat, model *deep learning* kini lebih diutamakan misalnya seperti model *Bidirectional Encoder Representations from Transformer* (BERT), di mana model ini merupakan salah satu model populer dalam lingkup klasifikasi teks dengan performa dan varian yang beragam yang juga menjadi pondasi atau model yang banyak digunakan dalam kebanyakan layanan Google. BERT merupakan model yang berbasis *transformer*, di mana model ini di desain dengan pendekatan bidireksional atau dua arah yang cocok digunakan untuk pengenalan teks [3].

Pemilihan data yang variatif dan seimbang, serta augmentasi data, penting untuk menghasilkan model yang optimal. Salah satu contoh augmentasi pada data teks misalnya penambahan atau penghapusan karakter pada teks [4]. Contoh augmentasi lainnya yaitu *Back Translation* dan *Synonym Replacement*.

Setelah membangun model ML, perlu dibuat antarmuka sederhana menggunakan Streamlit, sebuah *tools* berbasis python yang digunakan untuk mengembangkan aplikasi website untuk ML dan *data science*, Streamlit dashboard mendukung pengunggahan file lokal serta juga *request* data dalam database melalui *Application Programming Interface* (API) yang tersedia [5].

Model ML dan layanan sederhana yang dikembangkan diharapkan dapat mempermudah pengklasifikasian sertifikat untuk rekognisi pembelajaran lampau, dengan menyelaraskan sertifikat yang dimiliki dengan mata kuliah yang relevan. Hal ini bertujuan untuk mengurangi kemungkinan kesalahan manusia dalam proses klasifikasi manual serta mempercepat validasi keterampilan individu.

## 1.2 Rumusan Masalah

Melihat latar belakang yang telah dijabarkan sebelumnya, maka didapatkan perumusan masalah sebagai berikut:

- a. Bagaimana mengimplementasikan model BERT untuk melakukan klasifikasi sertifikat berdasarkan mata kuliah terkait?
- b. Bagaimana dampak augmentasi data terhadap hasil akurasi pelatihan pada model BERT?
- c. Bagaimana membangun layanan sederhana untuk memberikan hasil klasifikasi sertifikat yang dimiliki sesuai dengan mata kuliah terkait?

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dirumuskan, tujuan penelitian ini yaitu untuk mengetahui dampak penerapan augmentasi data terhadap akurasi dan kinerja model BERT dalam tugas klasifikasi sertifikat berbasis dokumen yang terintegrasi ke dalam layanan website sederhana.

#### 1.4 Manfaat Penelitian

Penelitian ini bertujuan mempermudah proses klasifikasi sertifikat dalam rangka rekognisi pembelajaran lampau, dengan menyelaraskan sertifikat yang dimiliki individu dengan mata kuliah relevan untuk mempercepat validasi keterampilan yang dimiliki. Dengan model ML BERT, penelitian ini mengembangkan layanan desain antarmuka berbasis Streamlit untuk mengklasifikasikan sertifikat yang diunggah sesuai mata kuliah terkait. Hasil klasifikasi ini diharapkan dapat dimanfaatkan institusi pendidikan atau perusahaan sebagai acuan rekognisi pembelajaran, sehingga memungkinkan integrasi keahlian non-kurikulum ke dalam program akademik atau kebutuhan industri.

#### 1.5 Batasan Masalah

Dalam penelitian ini dibutuhkan batasan masalah agar terdapat batas yang jelas serta lebih terarah. Batasan masalah sebagai berikut:

- a. Bahasa yang digunakan pada data sertifikat mencakup Bahasa Inggris dan Bahasa Indonesia.
- b. Augmentasi yang digunakan sebagai komparasi teknik augmentasi pada penelitian ini adalah *Character Insertion*, *Character Deletion*, *Back Translation* dan *Synonym Replacement*.
- c. Klasifikasi dibatasi pada tujuh kelas (mata kuliah): Machine Learning, Pemrograman Web, Desain Antarmuka, Jaringan Komputer, Aplikasi Game, Pemrograman Mobile, Manajemen proyek.
- d. Model yang digunakan adalah pre-trained BERT dengan dua macam konfigurasi yaitu base-uncased dan multilingual-uncased.

*Halaman ini sengaja dikosongkan*