

1. Pendahuluan

Latar Belakang

Tidak hanya sebagai tempat bersosialisasi, media sosial seperti Twitter juga dapat menjadi sebuah platform penyebaran informasi dengan berbagai macam format seperti teks, gambar, dan video. Informasi yang mengalir pada media sosial dapat membahas topik yang beragam variasinya dari aktivitas keseharian seorang pengguna, sampai berita peristiwa besar dalam kehidupan. Sosial media juga dapat bersaing dengan outlet berita lainnya sebagai sumber berita terkini bagi beberapa orang.

Sama halnya seperti informasi lainnya yang disebar, lembaga pendidikan seperti universitas juga memanfaatkan platform ini sebagai tempat penyebaran informasi. Salah satunya yaitu informasi mengenai beasiswa. Sehingga Twitter dapat menjadi salah satu tempat ideal yang digunakan untuk mengumpulkan data penelitian. Namun, kebanyakan pengguna Twitter menyebarkan informasi tersebut menggunakan format tersendiri. Hal ini menyebabkan data yang dikumpulkan menjadi tidak terstruktur dan dapat memperlambat usaha pengguna dalam mengumpulkan informasi penting. Melihat cepatnya informasi berlalu serta banyaknya kompetisi dalam pencarian kesempatan seperti beasiswa, efisiensi merupakan faktor penting dalam perlombaan meraih peluang yang terbatas tersebut.

Berdasarkan permasalahan yang diuraikan, Tugas Akhir ini mengusulkan untuk melakukan *Information Extraction* (IE) untuk mengambil informasi-informasi yang relevan dan mengubah data tidak terstruktur menjadi data terstruktur dengan format yang mudah dipahami. Ekstraksi informasi didefinisikan sebagai ekstraksi otomatis informasi tertentu dari teks bahasa alami. Lebih khusus lagi, IE secara otomatis memproses teks bahasa alami untuk mengekstrak informasi dari kelas entitas, hubungan, atau peristiwa tertentu[10].

Dalam penelitian ini, dilakukan implementasi ekstraksi informasi yang relevan dari kumpulan *tweet* bahasa Inggris mengenai topik beasiswa S3 yang dikirimkan oleh berbagai pihak pengguna Twitter. Implementasi dilakukan dengan melaksanakan tugas *Sequence Labelling* yaitu salah satu metode yang digunakan dalam *Natural Language Processing* (NLP). *Sequence Labelling* (SE) adalah jenis tugas pengenalan pola di cabang penting dari NLP. Pelaksanaan tugas *Sequence Labelling* dilakukan dengan membangun model yang menggunakan arsitektur *Bidirectional Long Short Term Memory Network* (BiLSTM) dan *Conditional Random Fields* (CRF).

BiLSTM-CRF digunakan karena BiLSTM dapat mengatasi *vanishing and exploding gradient problem* dan memberikan konteks informasi dua arah untuk mengatasi ketergantungan jarak jauh antar kata[12]. komponen CRF sering digunakan untuk mengatasi masalah berbasis sekuens dalam NLP. CRF memberikan informasi label tambahan pada tingkat kalimat[15]. Penggabungan CRF dengan BiLSTM juga menunjukkan peningkatan performansi dalam melakukan tugas *sequence labelling*[12].

Topik dan Batasannya

Berdasarkan uraian pada latar belakang, dapat diangkat permasalahan utama pada tugas akhir ini, yaitu bagaimana cara mengimplementasikan model ekstraksi informasi beasiswa pada *tweet* pengguna Twitter menggunakan *BiLSTM-CRF*. Adapun batasan dari permasalahan yang diangkat, yaitu:

1. Dataset yang digunakan merupakan informasi beasiswa S3 yang diperoleh dari *tweet* pengguna twitter.
2. Data menggunakan Bahasa Inggris
3. Data yang dimiliki menggunakan format skema BIO dimana labelnya terbagi menjadi:
 - a. *Field* : Bidang yang berkaitan dengan beasiswa
 - b. *Univ* : Universitas dikeluarkannya beasiswa tersebut
 - c. *Country* : Negara tempat dikeluarkannya beasiswa tersebut
 - d. *Link* : Tautan yang memiliki infor terkait dengan beasiswa yang dikeluarkan
 - e. *Deadline* : Batas waktu tersedianya pendaftaran beasiswa tersebut.
 - f. *Topic* : Topik yang berkaitan dengan beasiswa yang dikeluarkan
 - g. *Lab* : Lembaga lab yang berkaitan dengan beasiswa yang dikeluarkan

Tujuan

Tujuan dari pelaksanaan tugas akhir ini adalah mengimplementasikan model yang dapat melakukan ekstraksi informasi beasiswa S3 menggunakan arsitektur *BiLSTM-CRF* pada dataset Twitter. Hasil yang didapat kemudian dianalisis dan dilakukan evaluasi terhadap performansi model yang diimplementasikan.

Organisasi Tulisan

Penyusunan Laporan Tugas Akhir ini terbagi menjadi 5 bagian. Dimulai pada bagian pertama, terdapat pembahasan mengenai pendahuluan yang mencakup latar belakang, topik dan batasannya, serta tujuan pelaksanaan Tugas akhir ini. bagian kedua menjelaskan studi terkait dan penelitian lainnya yang berhubungan dengan topik pada Tugas akhir ini. Bagian ketiga menjelaskan arsitektur yang digunakan, yaitu *BiLSTM-CRF* dan alur kerja model yang dibangun. Bagian keempat menjelaskan hasil analisis dari keluaran model dan kinerjanya. Bagian kelima membahas kesimpulan yang dapat ditarik dari pelaksanaan Tugas Akhir ini.