

ABSTRAK

Implementasi Model XLM-RoBERTa Pada *Question Answering* Dataset SQuAD 2.0

Muhamad Rizal Firmansyah – 1187050064

Jurusan Teknik Informatika

Banyaknya informasi yang tersedia di internet berupa teks menyebabkan suatu masalah *information overload* muncul. Pengguna internet di era 4.0 mungkin merasa terlalu banyak informasi yang tersedia, sehingga sulit bagi mereka untuk memproses dan menyimpan pengetahuan secara efektif. Sistem *question answering* dapat membantu pembelajar di era 4.0 dengan cepat dan akurat mendapatkan informasi yang mereka butuhkan dari banyak informasi yang tersedia. Dengan memanfaatkan teknologi *natural language processing* khususnya *question answering* ekstraksi pengetahuan dari informasi yang berjumlah banyak dapat menjadi lebih cepat dan akurat. Metode yang digunakan untuk pembangunan model *question answering* yaitu CRISP-DM lalu model yang digunakan yaitu XLM-RoBERTa sebagai *pre-trained* model yang akan dibangun menjadi model *question answering*. Data yang digunakan untuk pelatihan adalah Stanford *Question Answering Dataset* atau SQuAD 2.0 versi ukuran kecil dan besar. Tujuan penelitian ini adalah untuk menerapkan dan mengukur kinerja model XLM - RoBERTa pada *Question Answering* pada dataset bahan ajar berupa data validasi SQuAD 2.0, dan data sejarah UIN dengan hasil yang diperoleh berhasil mendapatkan kinerja optimal nya yang mencapai rata-rata *recall* 71.3%, *precision* 70.1%, 68.4% dan *accuracy* 63.5%.

Kata kunci: CRISP-DM, *information overload*, *question answering*, SQuAD 2.0, XLM-RoBERTa.