

Abstrak

Alquran diturunkan dalam Bahasa Arab yang memiliki struktur bahasa kompleks, sistem penulisan yang unik, dan tata bahasa yang rumit sehingga menjadikannya sulit dipahami. Oleh karena itu, pemahaman dan interpretasi Alquran menjadi tujuan utama bagi umat Islam. Untuk memahami ajaran yang terkandung dalam Alquran, umat Islam membutuhkan pemahaman tentang entitas manusia yang ada di dalamnya. Namun, proses pelabelan entitas manusia dalam Alquran secara manual dapat menjadi tugas yang rumit dan menyebabkan kesalahan. Tujuan penelitian ini adalah memudahkan proses pelabelan entitas manusia dalam teks Alquran dengan membangun model yang memiliki performa yang baik.

RoBERTa merupakan model *Named Entity Recognition* (NER) berupa pengembangan dari BERT yang dilatih dengan metodologi pelatihan yang ditingkatkan. Penelitian ini fokus pada penggunaan model RoBERTa untuk mengidentifikasi entitas manusia dalam teks terjemahan Alquran Bahasa Indonesia. Masukan dari sistem ini berupa kalimat terjemahan Alquran yang kemudian diolah oleh model sehingga menghasilkan keluaran berupa label prediksi dari entitas kalimat tersebut. Model dibangun dengan memanfaatkan dataset dari korpus Tanzil Quran yang terdiri dari juz 1 hingga 6. Pra-pemrosesan data dilakukan dengan eksplorasi data dan *case folding*. Dataset dibagi menjadi data pelatihan (80%) dan data pengujian (20%). Model RoBERTa dilatih dengan *hyperparameter* berupa *epoch*, *learning rate*, dan *batch size*. Evaluasi dilakukan menggunakan perhitungan metrik *Precision*, *Recall*, dan *F-Score* pada data pengujian. Hasil evaluasi dari model RoBERTa yang dibangun memiliki nilai F-Score 52%. Nilai tersebut tidak lebih baik dibandingkan model BERT, yang menunjukkan bahwa model RoBERTa cenderung memiliki performa yang lebih buruk dalam mengidentifikasi entitas manusia pada teks terjemahan Alquran.

Kata kunci: entitas manusia, Alquran, Bahasa Indonesia, NER, RoBERTa