

### **Abstract**

*The Quran was revealed in Arabic, which has a complex linguistic structure, a unique writing system, and intricate grammar, making it challenging to understand. Therefore, understanding and interpreting the Quran is a primary goal for Muslims. To comprehend the teachings contained in the Quran, Muslims need an understanding of the human entities mentioned in it. However, manually labeling human entities in the Quran can be a complex task prone to errors. The aim of this research is to facilitate the process of labeling human entities in Quranic texts by building a model with good performance.*

*RoBERTa is a Named Entity Recognition (NER) model that is an extension of BERT, trained with enhanced training methodologies. This study focuses on the use of the RoBERTa model to identify human entities in the translated text of the Quran in Bahasa Indonesia. The input to this system consists of translated Quranic sentences, which are then processed by the model to generate output in the form of predicted labels for those sentence entities. The model is constructed by utilizing a dataset from the Tanzil Quran corpus, which comprises Juz 1 to 6. Data pre-processing is performed through data exploration and case folding. The dataset is divided into training data (80%) and testing data (20%). The RoBERTa model is trained with hyperparameters such as epochs, learning rate, and batch size. Evaluation is performed using metrics such as Precision, Recall, and F-Score on the testing data. The evaluation results of the constructed RoBERTa model show an F-Score value of 52%. This score is not better compared to the BERT model, indicating that the RoBERTa model tends to have inferior performance in identifying human entities in the translated text of the Quran.*

**Keyword:** *people entity, Quran, Indonesian, NER, RoBERTa*