ABSTRACT

The difficulty of managing administrative documents at St. John the Apostle Chapel, which is done manually, makes it prone to human error such as recording errors and document loss. This study aims to address these issues by processing the chapel's archive data in PDF and image formats, classifying them into three categories (permit letters, invitations, and loan agreements), and extracting important information from them. This integrated system employs Optical Character Recognition (OCR) technology using PyTesseract and Natural Language Processing (NLP) based on a fine-tuned IndoBERT model. Model testing results demonstrate very high performance, with the classification system achieving 93.6% accuracy on new test data, while the Named Entity Recognition (NER) model achieved an F1-Score of 0.989. User acceptance testing also demonstrated that the system is highly effective, successfully improving document management process efficiency by 69.6% based on time analysis and achieving a System Usability Scale (SUS) score of 77.5, indicating good usability. In conclusion, this system has proven to be functional, efficient, and well-received by users for enhancing the accuracy and efficiency of administrative document management.

Keywords: *Document Management, Optical Character Recognition, Natural Language Processing.*