

## OCR-Based Personal Loan Document Processing

In the evolving financial landscape, Optical Character Recognition (OCR) technology has revolutionized personal loan document processing. This report outlines the methodology, results, and conclusions of implementing an OCR-based system, emphasizing its impact on efficiency, accuracy, and user experience.

### Methodology

- **Technology Stack:** The system utilizes OCR powered by Tesseract, along with Python libraries like OpenCV for image processing and Streamlit for the user interface.
- **Data Preprocessing:** Images of loan documents are converted to grayscale, smoothed with Gaussian blur, and thresholded to enhance text visibility for OCR.
- **Text Extraction:** The Tesseract engine extracts text, which is processed to identify key fields such as the applicant's name, address, income, and loan amount.
- **Data Validation:** A thorough validation process checks for completeness and accuracy of the extracted data, ensuring integrity.
- **User Interface:** A user-friendly interface allows for easy document uploads and real-time feedback on data extraction, enhancing user engagement.

### Results

- **Efficiency Gains:** The OCR system significantly reduces manual data entry time, leading to faster loan approvals.
- **Accuracy Improvement:** Integration of AI with OCR enhances accuracy, minimizing human errors and ensuring reliable data processing.
- **User Experience:** The intuitive interface improves user confidence, providing immediate feedback and control over the data.
- **Validation Feedback:** Real-time notifications about discrepancies allow users to correct errors before submission, improving data integrity.

### Conclusions

- **Enhanced Processing Speed:** The automated system accelerates loan document processing, enabling quicker decision-making and improved customer satisfaction.
- **Cost Reduction:** Reduced reliance on manual labor leads to significant operational savings for financial institutions.
- **Scalability:** The cloud-based architecture efficiently handles varying document volumes, allowing institutions to adapt to peak times without extra costs.

### Rationale for Selected Approach

- **User -Centric Design:** Streamlit was selected for its effectiveness in creating interactive applications, promoting user engagement.
- **Focus on Compliance and Security:** Emphasis on data validation ensures regulatory compliance and high data integrity standards.