BFSI- OCR OF BANK STATEMENTS

Milestone 2 Report

PREPARED BY:
Jasneet Arora

Table of Contents

1. Introduction

The BFSI (Banking, Financial Services, and Insurance) OCR (Optical Character Recognition) project aims to develop a comprehensive optical character recognition system for automated bank statement analysis, integrating advanced text extraction technologies. Milestone 2 focuses on implementing a sophisticated OCR comparator that leverages multiple extraction techniques to address the critical challenge of accurately extracting textual information from complex financial documents. By developing a flexible framework that combines artificial intelligence and traditional OCR methods, the milestone seeks to establish a robust foundation for automated financial document processing.

2. Objectives

Milestone 2 focuses on developing an OCR extraction framework for reliably processing financial documents. The system will integrate various text recognition technologies, compare OCR methods, and create a flexible approach for diverse document formats. The goal is to build a tool that accurately extracts text from bank statements, enabling advanced financial data analysis.

3. Methodology

The research methodology adopts a systematic approach to OCR technology evaluation. An extensive investigation identifies and selects the most appropriate optical character recognition techniques suitable for financial document processing. Three primary OCR technologies are carefully selected: LlamaOCR, an advanced AI model; EasyOCR, a traditional machine learning technique; and Tesseract OCR, a robust open-source solution. This multi-modal approach ensures comprehensive and versatile text extraction capabilities.

4. Technical Implementation

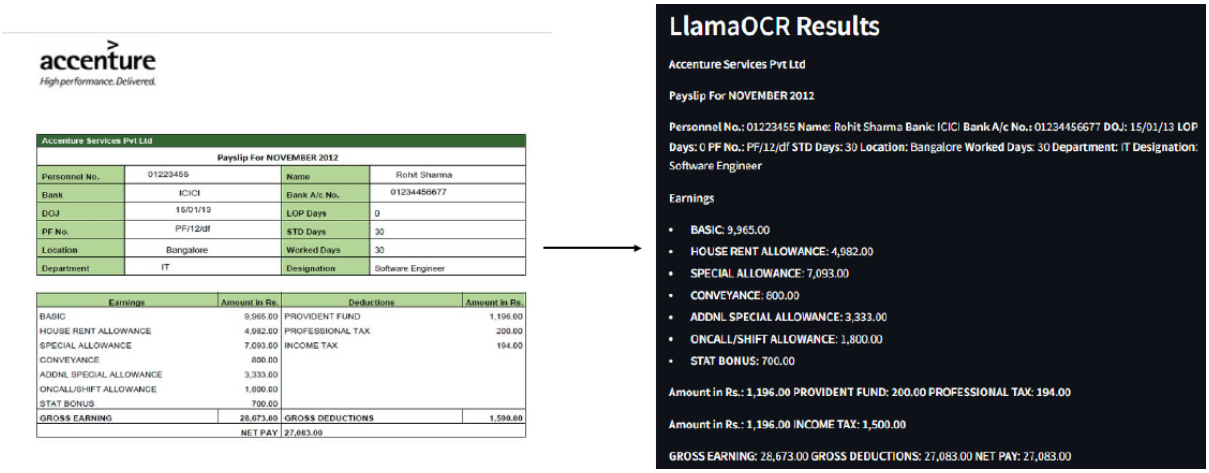
4.1 Overview of OCR Techniques

Milestone 2 leverages three distinct OCR technologies, carefully selected to provide comprehensive text extraction capabilities for bank statement analysis. By integrating advanced AI-powered solutions, traditional machine learning techniques, and robust open-source OCR engines, the milestone aims to develop a sophisticated document processing system.

4.2 LlamaOCR Implementation

LlamaOCR represents the most advanced approach in the Milestone 2 OCR framework, utilizing the meta-llama/Llama-3.2-11B-Vision-Instruct-Turbo model. The LlamaOCR processor was configured to extract text with a focus on comprehensive content understanding. The implementation goes beyond traditional text extraction by providing advanced image analysis capabilities.

The model is specifically configured to extract text from financial documents, with a custom prompt designed to capture precise textual information. Base64 image encoding ensures seamless processing and robust error handling manages various document scenarios. The extraction process leverages the model's advanced language understanding to interpret complex financial document layouts.

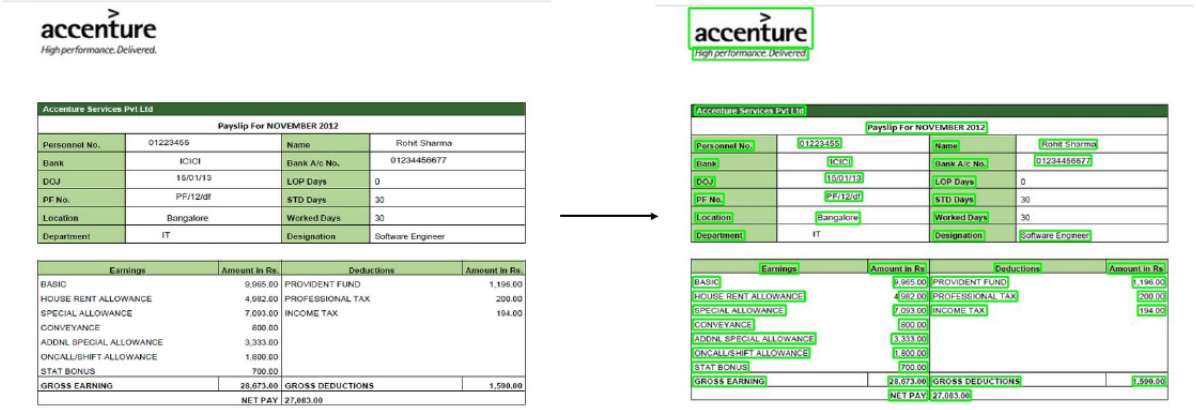


4.3 EasyOCR Implementation

EasyOCR serves as the secondary OCR technique, providing a robust machine learning-based approach to text extraction. For Milestone 2's bank statement analysis, the EasyOCR processor was carefully configured to optimize performance for financial document processing.

The implementation includes fine-tuned text thresholds to improve accuracy in extracting financial text. Specific parameters were set for text threshold, low text threshold, and link threshold to balance comprehensive text capture and precision. The canvas size and magnification ratio were adjusted to handle various bank statement formats effectively.

The processor prioritizes English language processing, with GPU acceleration enabled to ensure rapid and efficient text extraction. Advanced bounding box extraction provides precise location information for each extracted text element, crucial for detailed financial document analysis.

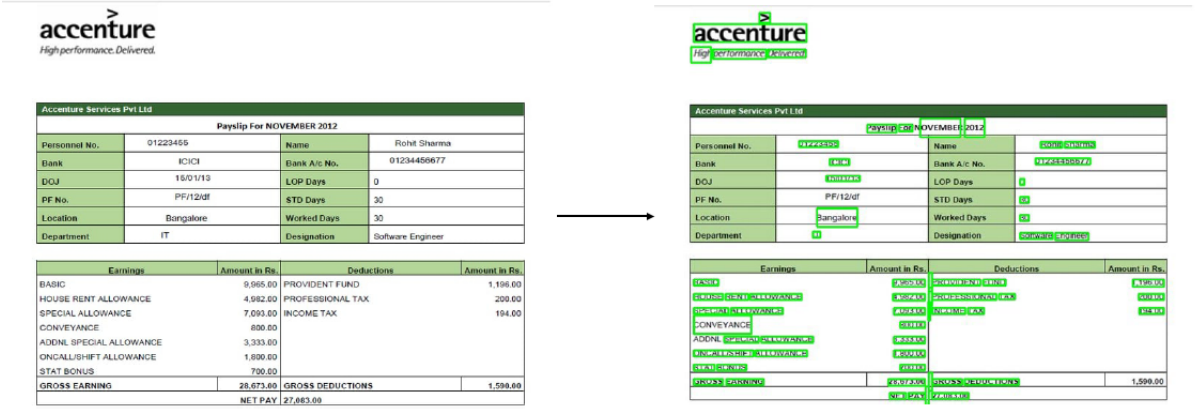


4.4 Tesseract OCR Implementation

Tesseract OCR provides the third layer of text extraction, offering a reliable open-source solution for document processing. For Milestone 2's bank statement analysis, Tesseract was configured with specific parameters to enhance its effectiveness in financial document analysis.

The Page Segmentation Mode (PSM) and OCR Engine Mode (OEM) were carefully selected to optimize text extraction. A minimum confidence threshold filters out low-reliability text extractions, ensuring only high-quality text is retained. The implementation focuses on English language processing, with a custom configuration designed to handle the specific layouts and text structures common in bank statements.

The Tesseract processor provides detailed extraction metadata, including precise bounding box coordinates and confidence scores for each extracted text element. This approach allows for comprehensive analysis and comparison of text extraction results across different OCR techniques.



4.5 Integration Approach

The OCR comparator integrates these three technologies through a unified processing framework. This approach enables simultaneous processing using multiple OCR techniques, providing a comprehensive and comparative analysis of text extraction results. The implementation offers flexible hyperparameter configuration and visualization of extraction outcomes, ensuring a robust and adaptable solution for bank statement analysis.

By combining LlamaOCR's advanced AI capabilities, EasyOCR's machine learning approach, and Tesseract's reliable open-source processing, Milestone 2 creates a multi-modal OCR system specifically designed for extracting and analysing financial document information.

5. Results and Deliverables

Milestone 2 successfully demonstrated the effectiveness of multiple OCR techniques in text extraction. Each OCR method showed unique strengths, with LlamaOCR providing advanced semantic understanding, EasyOCR exhibiting robust text detection, and Tesseract OCR delivering precise character recognition. The implementation resulted in a comprehensive comparative analysis framework with a Streamlit-based interface that allows dynamic technique selection and result visualization.

6. Challenges and Solutions

During implementation, several technical challenges were identified and addressed. Diverse financial document formats presented significant complexity, requiring adaptive preprocessing techniques to normalize documents from multiple sources. Varying image qualities, including inconsistent resolutions and scanning conditions, were mitigated through advanced image enhancement algorithms that improved text extraction reliability.

The integration of multiple OCR technologies demanded a sophisticated approach. A flexible framework was developed to dynamically select and combine extraction techniques based on specific document characteristics, ensuring optimal text recognition across different document types.

7. Next Steps

The next phase of the project will focus on enhancing the analysis capabilities by incorporating semantic analysis, anomaly detection, and predictive modeling. We will also strive to improve the user experience by designing a more intuitive interface and seamless integration with popular financial tools. Additionally, we will continue to optimize the OCR models and explore cloud deployment for scalability and accessibility.

8. Appendices

Appendix A: Technical Architecture

The project leverages a multi-modal approach, combining AI, machine learning, and traditional OCR techniques. The architecture comprises the following key components:

- Data Ingestion
- Pre-processing
- OCR Engine
- Post-processing
- Data Analysis

- User Interface

Appendix B: Technical Specifications

- **Hardware Requirements:**
    - CPU: Intel Core i5 or equivalent
    - GPU: NVIDIA GeForce GTX 1060 or equivalent (optional for accelerated processing)
    - RAM: 8GB or more
    - Storage: 256GB SSD or more
- **Software Requirements:**
    - Python 3.x
    - TensorFlow or PyTorch
    - OpenCV
    - Tesseract OCR
    - EasyOCR
    - Streamlit (for UI)
    - NLTK (Natural Language Toolkit)
    - Other relevant libraries and frameworks