Vivekanand Education Society's Institute of Technology

Department of Computer Engineering

Group No.: 13

Date : 02/08/2024

Project Synopsis (2024-25) - Sem V

Document Validation System

Under the guidance of

Dr.Prashant Kanade
Asst. Professor, Computer Engineering Department

Taniya Vallecha(2022.taniya.vallecha@ves.ac.in)
Nimish Chug(2022.nimish.chug@ves.ac.in)
Meet Mattani(2022.meet.mattani@ves.ac.in)
Darshan Khapekar(d2022.darshan.khapekar@ves.ac.in)

Abstract:

This project focuses on the development of an automated document verification system designed to ensure data integrity and streamline the application review process. The system allows users to upload documents alongside a form submission, and it uses advanced Intelligent Document Processing (IDP) techniques to verify that the information in the documents matches the data provided by the applicant. In case of discrepancies, the system triggers alerts, prompting the applicant to correct or resubmit their documents. This approach not only speeds up the verification process but also significantly reduces the likelihood of human error, making the application process more efficient and reliable.

Introduction:

In today's digital age, the need for efficient and accurate document verification systems has become increasingly critical across various sectors. Traditional manual methods of verifying documents are time-consuming, prone to errors, and often lead to delays in processing applications. This project addresses these challenges by developing a sample application that automates the document verification process. By leveraging Intelligent Document Processing (IDP) techniques, the system extracts information from uploaded documents, compares it with the data provided by the applicant, and immediately flags any discrepancies.

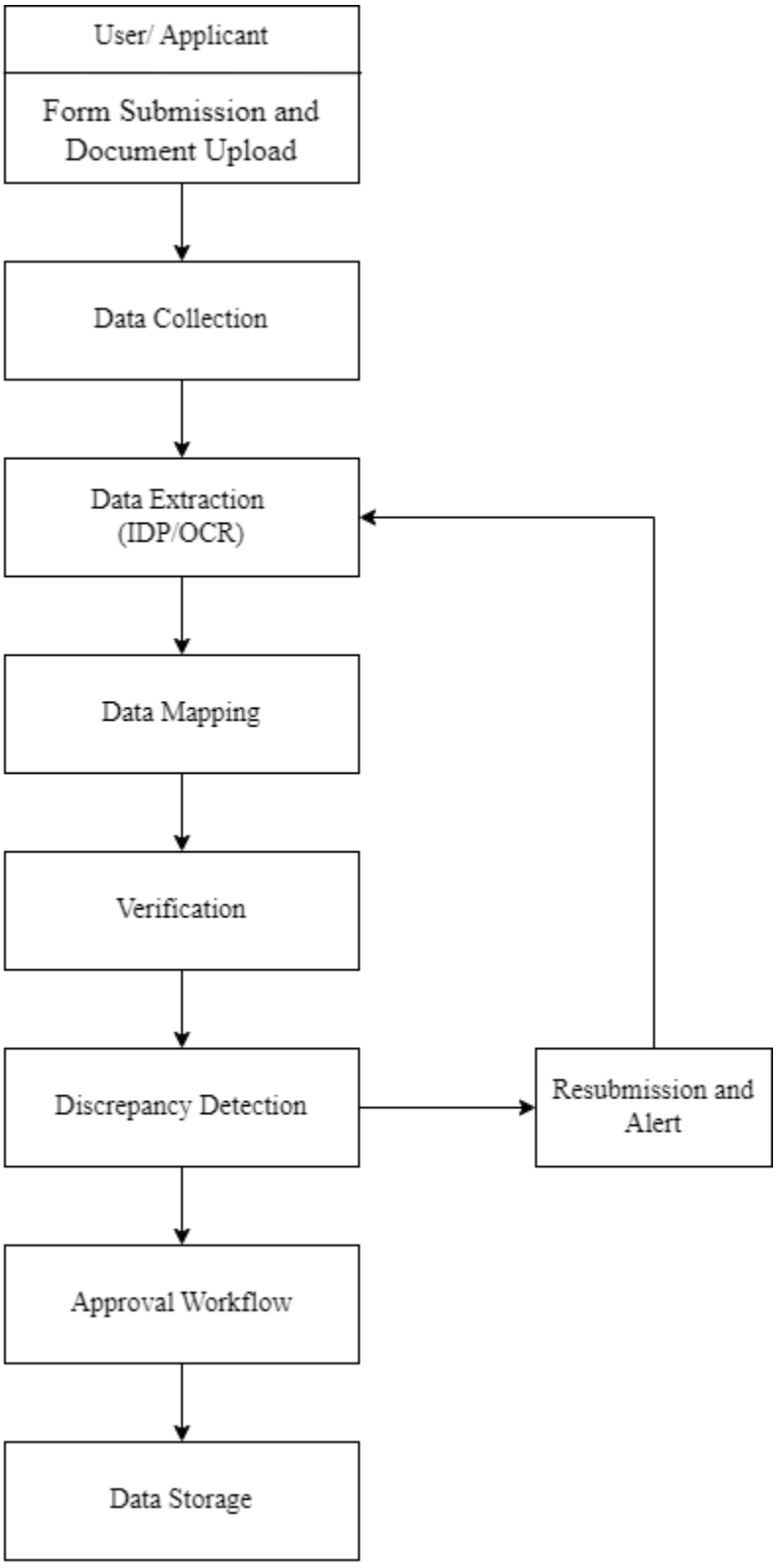
Problem Statement:

Manual document verification is a resource-intensive process that often leads to delays, inaccuracies, and inconsistencies. The primary challenges include the time required to verify documents, the potential for human error, and the difficulty in managing large volumes of applications. Additionally, in scenarios where documents are submitted in different formats , the verification process becomes even more complex. These issues can result in prolonged application processing times and may compromise the integrity of the data collected.

Proposed Solution:

The proposed solution is an automated document verification system that integrates a user-friendly application form with a robust document upload and verification mechanism. The system employs Machine Learning, Deep Learning, and Natural Language Processing (NLP) techniques to extract and analyze information from documents. It compares the extracted data with the information provided by the applicant and triggers alerts in case of any mismatch. This ensures that only accurate and consistent data is entered into the system, thereby improving the overall efficiency and reliability of the application review process.

Block Diagram:



Methodology:

- **User/Applicant - Form Submission and Document Upload:**

  The applicant fills out the online form and uploads the required documents (e.g., educational certificates, ID proofs).

- **Data Collection:**

  The system collects the submitted data and documents for further processing.

- **Data Extraction:**

  Using Intelligent Document Processing (IDP) and Optical Character Recognition (OCR) techniques, the system extracts textual information from the uploaded documents.

- **Data Mapping:**

  The extracted data is mapped to the corresponding fields provided in the form by the applicant.

- **Verification:**

  The system verifies the accuracy of the data by comparing the mapped information from the documents against the details entered in the form.

- **Discrepancy Detection:**

  If there are any mismatches or inconsistencies between the form data and the extracted document data, the system detects these discrepancies.

- **Resubmission and Alert (in Case of Discrepancy):**

  In case of detected discrepancies, the system triggers an alert to the applicant, prompting them to recheck the entered data or resubmit the correct documents.

- **Approval Workflow (if No Discrepancy):**

  If no discrepancies are found, the system moves the application forward into the approval workflow.

- **Data Storage:**

  The verified data is securely stored in the system's database for future reference and reporting.

Hardware, Software, and Tools Requirements:

Hardware Requirements:
Processor: Quad-core CPU or higher (Intel i5/i7 or AMD Ryzen 5/7).
RAM: Minimum 16 GB (32 GB recommended).
Storage: SSD with at least 512 GB (NVMe SSD preferred).
GPU (Optional): For accelerated machine learning tasks, a mid-range GPU like NVIDIA GTX 1660 or higher.

Software Requirements:
Operating System: Windows, Linux, or macOS
Database: Firebase
Programming Languages:
Python: For implementing Intelligent Document Processing (IDP) using libraries like Tesseract OCR, OpenCV, and NLP tools.
Frameworks and Libraries:
Flask/Django: For backend API development
TensorFlow/PyTorch: For implementing any machine learning models required for document classification or verification.

Proposed Evaluation Measures:

- **Accuracy of Data Matching:**

  Ensure the system accurately matches the information in the uploaded documents with the details provided by the applicant.

- **Effectiveness in Mismatch Detection:**

  Evaluate how effectively the system detects discrepancies between the applicant's entered data and the data extracted from the documents.

- **Document Processing Reliability:**

  Assess the system's ability to consistently extract and compare data from different document types and formats.

- **Scalability Testing:**

  Determine how well the system handles increasing volumes of applications without compromising its accuracy.

Conclusion:

The development of an automated document validation system using Intelligent Document Processing techniques offers significant advantages in terms of efficiency, accuracy, and reliability. By automating the verification process, the system reduces the time and resources required to process applications while minimizing the risk of errors and inconsistencies. The alert mechanism enhances user experience and ensures that data integrity is maintained throughout the process. This project demonstrates the potential of advanced technologies to transform traditional manual processes into streamlined, automated systems that better serve organizations and applicants alike.

References:

- Karan Ingale, Rohan Konde, Sharayu Khachane, Hrutika Suryawanshi, Jagdish Kapadnis. *Enhancing Document Verification with Digital Signature and OCR Algorithm*. International Research Journal of Modernization in Engineering Technology and Science, Volume 05, Issue 11, November 2023.
- Aryan Salge, Srushti Shindkar, Sanika Malve, Shreyas Dabhikar, Prof. Sharada Desai. *Document Verification Using OCR* , https://ymerdigital.com.
- S. Mthethwa, N. Dlamini. *Verifying the Integrity of Hard Copy Document Using OCR*. researchspace.csir.co.za.