OCR based loan Application Processing

Submitted By Rishabh (DT20234378792)

Problem:

Banks handle numerous personal loan applications, which often require manual data entry and verification from various documents such as ID proofs, income statements, and bank statements. This process is time-consuming and prone to errors. Optical Character Recognition (OCR) technology can automate the extraction of relevant information from these documents, improving efficiency and accuracy.

Objective:

Develop an OCR-based solution to automatically extract and process information from personal loan application documents. The system should identify key fields such as applicant name, address, income details, and loan amount

Approach:

- If pdf convert it into image to send to ocr
- If image upload page1 and page2 mandatory
- Use Ocr for text detection
- Use regex for better text extraction from the ocr result
- Crop photo and sign of applicant from form based page height and width configuration
- The result will show in the form of a form to modify in case of issue in data extraction
- Then user confirm the showed data and submit the form

Methodology/Algorithm:

```
option=radioButton('pdf', 'image')

If option==pdf:

        Upload pdf

        Img1,img2=covertToImage(uploaded pdf)

        OcrData=imageOcrResult(img1,img2)

        Applicant_photo=getphotofromImg(img1)

        Applicant_sign=getSignfromImg(img2)

Else:

        Upload page1 as img1 and page2 as img2

        ocrData=imageOcrResult(img1,img2)
```

Applicant_photo=getphotofromImg(img1)

Applicant_sign=getSignfromImg(img2)

If ocrData:

Show all data in a form

Show applicant_photo and applicant_image and option to modify them

Checkbox=checkbox(text= "I have carefully reviewed the application and all data is correct.")

If checkbox:

If submit button pressed:

Print("Application submitted successfully")

Show applicant data in json

Note**

imageOcrResult(img1,img2) function: return text of image and to parse data by field regex is used

getphotofromImg(img) : The position of the applicant's photo is located by calculating its position as a percentage of the image's width and height, relative to the top-left corner. The photo is then cropped from the image based on this calculated region.
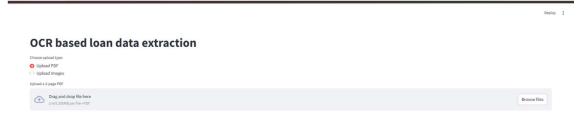
getSignfromImg(img): Similarly, the applicant's signature is cropped from the second page by calculating its position as a percentage of the image's width and height, relative to the top-left corner.
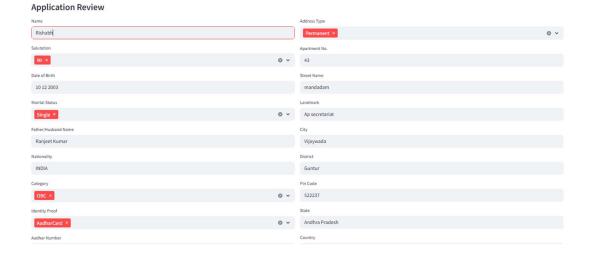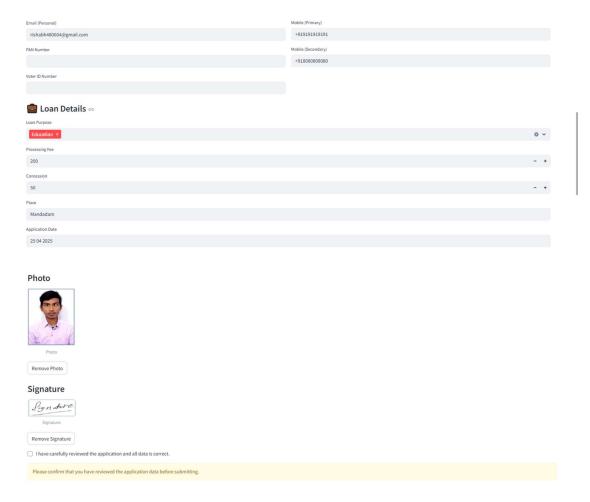
Screenshot:

Application Form

SBI Personal Loan FORM

SECTION-A (PERSONAL & KYC)

Name : Rishabh

Salutation : Y Mr _ Mrs _ Dr. _ Other

Date Of Birth : 10 12 2003

Martial Status : Y Single _ Married

Name of Father/Husband : Ranjeet Kumar

Nationality : INDIA   Category : _ SC _ ST Y OBC _ General

Identity Proof: Y AadharCard ___ PANCard ___ VoterIDCard

AADHAR NO : 1234 5678 9012

PAN NO :

VoterIDCard:

Address : Y Permanent ___ Correspondence

    Apartment No. or Name : 43

    Street Name or Area : mandadam

    Landmark : Ap secretariat

    City : Vijaywada   District : Guntur   Pin Code : 522237

    State : Andhra Pradesh   Country : India

Mobile (Primary): +919191919191   Mobile(Secondary) : +918080808080

Email(Personal) : rishabh480604@gmail.com

SECTION-B(Loan Details)

Scheme Name : Y Xpress credit ___ Other

Loan Details:

Loan Amount :   400000   Rate of Interest (%): 12

Loan Purpose : Y Education ____ Marriage ___ Medical ___ Other(specify)

Processing Fee : 200   Concession : 50

Declaration

I/We certify that the information provided by me/us in this application form is true and correct in all respects and State Bank of India is entitled to verify this directly or through any third party agent. I/We confirm that the attached copies of financials/Bank Statements etc., are submitted by me/us against my/our loan application and certify that these are true copies. I/We further acknowledge the Bank's right to seek any information from any other source in this regard. I/We understand that all of the above-mentioned information shall form the basis of any facility that the Bank may decide to grant to me/us at its sole discretion.

Place : Mandadam
Date : 25 04 2025

Signature

Signature of Applicant

Main Page

Deploy ⋮

OCR based loan data extraction

Choose upload type:
- ● Upload PDF
- ○ Upload Images

Upload a 2-page PDF

☁ Drag and drop file here
Limit 200MB per file • PDF

Browse files

After Uploading Application Form

Application Review

Name	Address Type
Rishabh	Permanent ✕

Salutation	Apartment No.
Mr ✕	43

Date of Birth	Street Name
10 12 2003	mandadam

Marital Status	Landmark
Single ✕	Ap secretariat

Father/Husband Name	City
Ranjeet Kumar	Vijaywada

Nationality	District
INDIA	Guntur

Category	Pin Code
OBC ✕	522237

Identity Proof	State
AadharCard ✕	Andhra Pradesh

Aadhar Number	Country

Email (Personal)	Mobile (Primary)
rishabh480604@gmail.com	+919191919191

PAN Number	Mobile (Secondary)
	+918080808080

Voter ID Number	

💼 Loan Details ∞

Loan Purpose

Education ×   ⚙ ⌄

Processing Fee

200   − +

Concession

50   − +

Place

Mandadam

Application Date

25 04 2025

Photo

Photo

Remove Photo

Signature

Signature

Remove Signature

☐ I have carefully reviewed the application and all data is correct.

Please confirm that you have reviewed the application data before submitting.

After Submission (json data):

Application Submitted Successfully!

```
▼ { 🗐
    "name" : "Rishabh"
  ▼ "salutation" : [
      0 : "Mr"
    ]
    "dob" : "10 12 2003"
  ▼ "marital_status" : [
      0 : "Single"
    ]
    "father_husband_name" : "Ranjeet Kumar"
    "nationality" : "INDIA" 🗐
  ▼ "category" : [
      0 : "OBC"
    ]
  ▼ "identity_proof" : [
      0 : "AadharCard"
    ]
    "aadhar_number" : "1234 5678 9012"
    "pan_number" : NULL
    "voter_id" : NULL
  ▼ "address_type" : [
      0 : "Permanent"
    ]
    "apartment_no" : "43"
    "street_name" : "mandadam"
    "landmark" : "An secretariat"
```

Conclusion:

In this project, we successfully built a system that can automatically extract important fields from personal loan application documents. It works for both PDF and image uploads — if a PDF is uploaded, it is first converted into images for further processing.

Using OCR combined with regex patterns, we were able to extract key information like applicant name, address, Aadhar number, PAN, voter ID, and category.

We also extracted the applicant's photo and signature by cropping the form based on the percentage of width and height relative to the page, making the extraction reliable across different document sizes.

The extracted data is displayed in a user-friendly form for manual review and correction before final submission.

Overall, the system significantly reduces manual effort, improves accuracy, and speeds up the loan processing workflow.

Why This Approach Was Selected:

This approach was chosen because it is simple, effective, and flexible. By using OCR and regex-based extraction, we avoided the need for complex AI models that require heavy training and data collection.

The system supports both **PDF and image inputs**, making it versatile for real-world scenarios where documents can come in different formats.

Using percentage-based cropping ensures that the photo and signature extraction works even if the document resolution changes.

Additionally, allowing manual correction after OCR helps tackle any mistakes due to poor scan quality or handwriting.

Overall, this method was selected because it balances **automation**, **accuracy**, and **ease of use**, making it a practical solution for real-world banking environments.