



OPTIPARSE



Pool NAWABS





COMPONENTS



01

SERVER and BACKEND

We utilized Azure's B2ms virtual machine (2 vCPU, 8 GB RAM), leveraging free credits for high-compute performance in server setup and backend integration.

02

FRONTEND

The app includes key features such as a Home Screen, Add Transaction, Search Page, and dynamic Charts, all optimized with ObjectBox for fast local storage and retrieval.

03

OCR

We implemented PaddleOCR's quantized PP-OCRv3 model with knowledge distillation to achieve efficient and accurate optical character recognition.

Api Specifications

Built with django, this api can be accessed via the IP:

<http://52.140.76.58:8000/api/>, or the DNS address

<http://takneeknawabs.centralindia.cloudapp.azure.com:8000/api>

- Methods :
- /api/ : a simple get request to check system health
 - /api/upload : to upload the image on the server and returns its respective index
 - /api/detect : returns the detected data by the model in form of JSON

The server has been dockerized and details of its setup has been mentioned in our doc

OCR & LLM

OCR - PaddleOCR with PP-OCRv3

- Framework: Leveraged PaddleOCR, based on Baidu's PaddlePaddle deep learning suite.
- Model: Utilized PP-OCRv3-slim, a quantized model employing knowledge distillation for reduced size without sacrificing accuracy.
- Accuracy: Achieved ~85% accuracy on key fields like Invoice ID, Total Amount, Vendor Name, and Date.
- Performance: Inference time averages 0.2s/image on a Colab T4 GPU.

PP-OCRv3 Pipeline

- Text Detection: LK-PAN for larger text detection, DML for mutual learning, and RSE-FPN for enhanced feature extraction.
- Text Recognition: SVTR-LCNet for speed and efficiency, using Guided Training and data augmentation (TextConAug, UIM) to enhance model accuracy.

LLM Integration with Mistral 7B

- Model Used: Mistral 7B Instruct v0.2, hosted on Huggingface, chosen over Llama 2-7B after performance comparisons.
- Task: Extract key financial information from OCR output and format it into JSON.


Prompt for LLM:

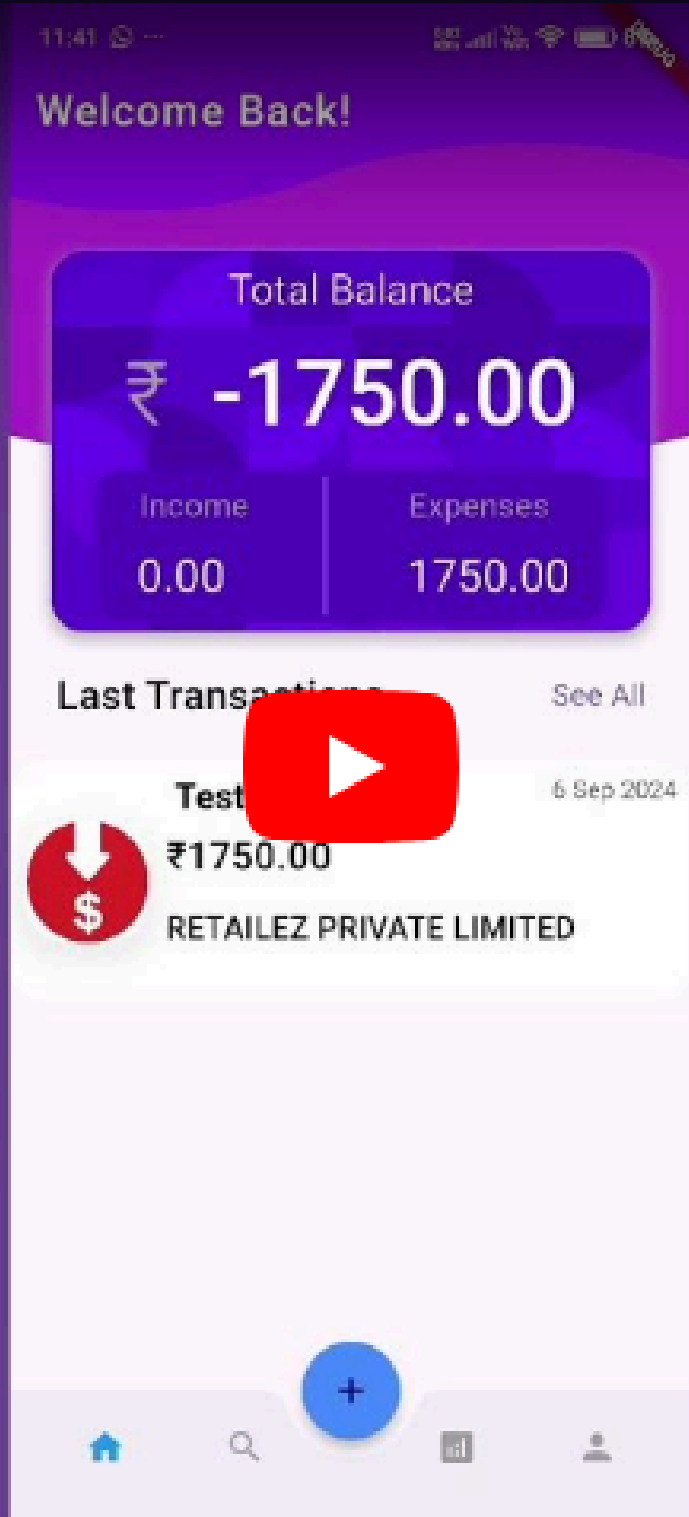
- Role: The LLM acts as a Financial Document JSON Parser.
- Output Format: JSON

```
{  
  "date": DD/MM/YYYY,  
  "invoice_id": string,  
  "total_amount": float,  
  "vendors_name": string,  
  "buyer_name": string  
}
```

Frontend Application

<https://www.youtube.com/watch?v=NRmoYbZmuQw>

Optiparse Function video



11:41

Welcome Back!

Total Balance

₹ -1750.00

| Income | Expenses |
|--------|----------|
| 0.00 | 1750.00 |

Last Transactions


Test

₹1750.00

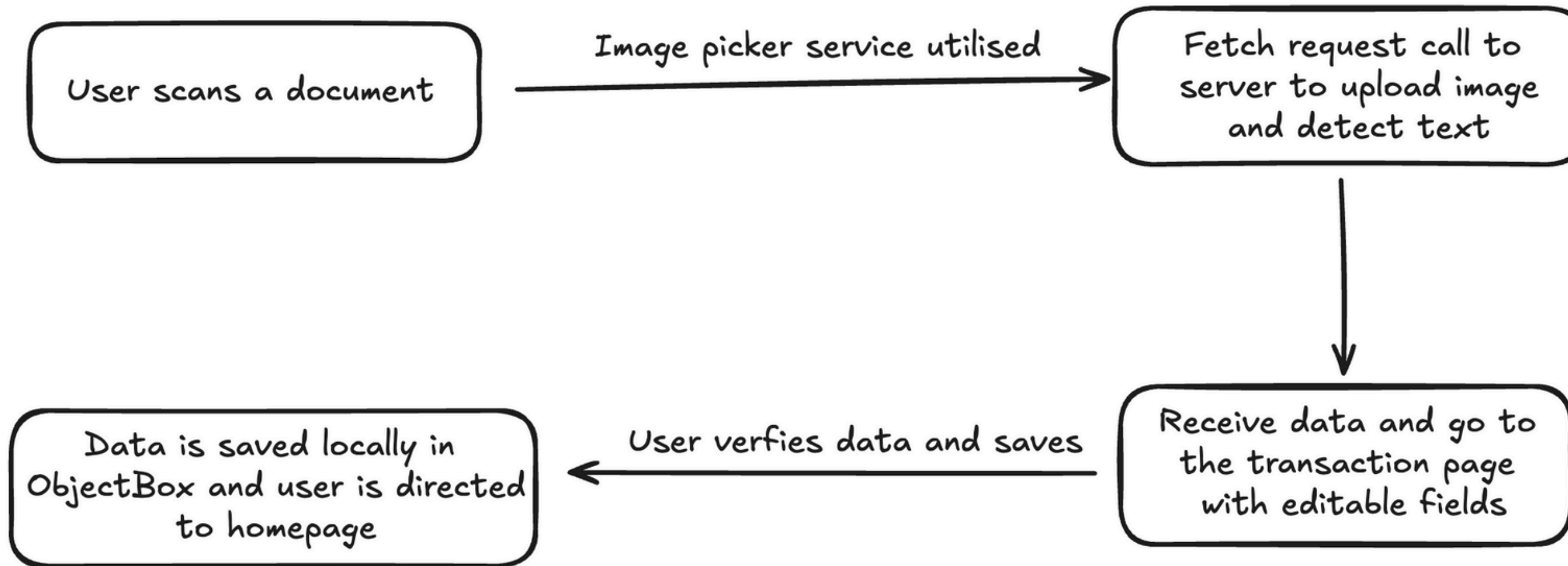
RETAILEZ PRIVATE LIMITED

6 Sep 2024

See All

Watch on  YouTube

Copy link



Flow of OCR

Thank you