

Paperless.AI

A revolutionary OCR platform that digitizes paper-based documents, transforming banking processes into efficient and eco-friendly operations

1 PROBLEMS

Manual invoice processing in the banking sector facing such challenges:

- High volume of invoices
- Variations in invoice formats
- Tedious manual data entry and verification processes



2 PROPOSED SOLUTION

A web app that allows user to scan all invoices in no time with:

- Automation in invoice processing
- Reducing human error
- Speed up financial operation
- Cost savings and compliance



3 DATASET

Invoice Images Dataset:

- <https://zenodo.org/record/6371710>
- <https://www.kaggle.com/datasets/holtskinner/invoices-document-ai>



4 IMPLEMENTATION

- **Preprocessing:** Enhance image quality, convert to OCR format
- **Feature Engineering:** Text localization & text segmentation
- **Modeling:** Pretrained Tesseract on invoice images
- **Evaluation:** Low Character Error Rate (CER) measures
- **Deployment:** Streamlit platform



5 EXAMPLE

Invoice	
DATE	INVOICE #
1/16/2018	118708
01/16/2018	118708

6 PROTOTYPE

