
Documentation for Fake Invoice Detection

Using Tesseract OCR and Computer Vision

Maha Qaiser

22i-2348

Description

A simple web application that verifies whether an invoice is original or fake:

- If the layout and fields match expected patterns -> Invoice is marked as Original
- If the layout or mismatches are detected -> Invoice is marked as Fake

Steps to Use the Web Application

Run: python app.py

Go to: <http://127.0.0.1:5000/> then upload an invoice image and click 'Analyze Invoice'

Flow of Web Application

Step 1: Upload Image

The user clicks the 'Choose File' button and uploads an image

Step 2: Click to Analyze

The user clicks the 'Analyze Invoice' button and the results are displayed

Solution

- Matches input image with template image using computer vision
- Extracts key fields from an invoice image using Tesseract OCR and regular expressions
- Classifies the invoice as 'Original' or 'Fake' based on the presence of certain fields

Assumption: Correct format of invoice matches the images of invoices from <https://www.kaggle.com/datasets/osamahosamabdellatif/high-quality-invoice-images-for-ocr>

Step 1: Load the invoice image

Step 2: Match input image with template image

Step 3: Use Tesseract OCR to extract all visible text from the image

Step 4: Use regular expressions to extract important fields:

- Invoice Number: Searches for a pattern like "Invoice no: 123456"
- Date of Issue: Tries to match "Date of issue: dd/mm/yyyy"
- If not found, looks for the word "Total" and then checks two lines below for a date
- IBAN: Looks for a valid alphanumeric IBAN string
- Tax IDs: Searches for two Tax IDs (format: 123-45-6789)
 - If two or more Tax IDs are found:
 - The first one is stored as the Seller's Tax ID
 - The second one is stored as the Client's Tax ID
 - Else both are marked as "not found"

Step 5: Classify the invoice:

- 'Fake Invoice' -> If any required field is missing or marked "not found"
- Otherwise, 'Original Invoice'

Test case 2

Fake Invoice Input:



INVOICE

Date Issued:
01 January 2030

Invoice No:
01234

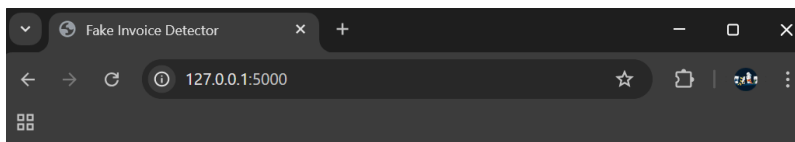
Issued to:
Sacha Dubois
123 Anywhere St., Any City,
ST 12345

NO	DESCRIPTION	QTY	PRICE	SUBTOTAL
1	Item 1	1	\$ 100	\$ 200
2	Item 2	2	\$ 50	\$ 100
GRAND TOTAL				\$ 300

Note:
Bank Name: Rimberio
Account No: 0123 4567 8901

Claudia
Finance Manager

Output:



Fake Invoice Detector

No file chosen

Extracted Fields:

- **Invoice No:** 123
- **Date of Issue:** not found
- **Seller Tax ID:** not found
- **Client Tax ID:** not found
- **IBAN:** not found

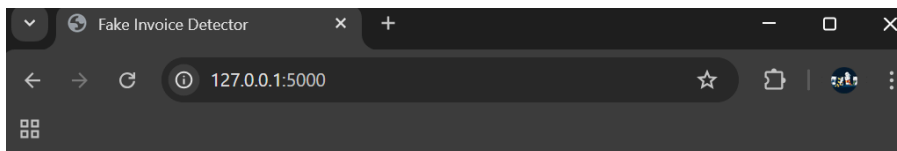
Fake Invoice

Test case 3

Fake Invoice Input:



Output:



Fake Invoice Detector

Choose file No file chosen

Analyze Invoice

Fake Invoice (layout mismatch)