
Documentation for Fake Invoice Detection

Maha Qaiser

22i-2348

Description

A simple web application that verifies whether an invoice is original or fake using computer vision techniques.

- 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

- 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 and convert it to grayscale to improve OCR accuracy

Step 2: Use Tesseract OCR to extract all visible text from the image

Step 3: 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 4: Classify the invoice:

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

Example Results

Test case 1

Original Invoice Input:

Invoice no: 40394912

Date of issue: 03/23/2020

Seller:
Mccullough-Bradley
457 Vargas Brook Suite 967
East James, MA 63095

Tax Id: 905-94-2945
IBAN: GB76GUQ007697012433171

Client:
Cross-Estrada
7297 Smith Roads
Lake Sherri, TN 88469

Tax Id: 915-81-8637

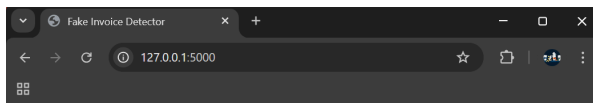
ITEMS

No.	Description	Qty	UM	Net price	Net worth	VAT [%]	Gross worth
1.	Microsoft Xbox One X 1TB 4k Ultra HD Console black controller works great 1787	5,00	each	249,00	1 245,00	10%	1 369,50

SUMMARY

	VAT [%]	Net worth	VAT	Gross worth
	10%	1 245,00	124,50	1 369,50
Total		\$ 1 245,00	\$ 124,50	\$ 1 369,50

Output:



Fake Invoice Detector

No file chosen

Extracted Fields:

- **Invoice No:** 40394912
- **Date of Issue:** 03/23/2020
- **Seller Tax ID:** 905-94-2945
- **Client Tax ID:** 915-81-8637
- **IBAN:** GB76GUQ007697012433171

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

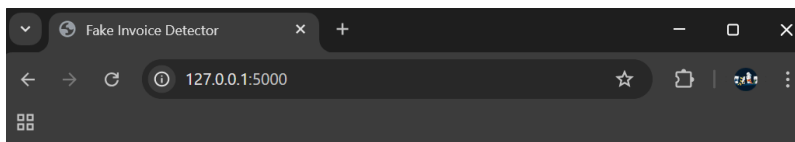
NO	DESCRIPTION	QTY	PRICE	SUBTOTAL
1	Item 1	1	\$ 100	\$ 200
2	Item 2	2	\$ 50	\$ 100
GRAND TOTAL				\$ 300

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

Claudia
Finance Manager



Output:



Fake Invoice Detector

Choose file No file chosen

Analyze Invoice

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

Original Invoice Input:

Invoice no: 51295021

Date of issue: 01/05/2017

Seller:
Sims-Olson
USS Bell
FPO AP 94327

Tax Id: 922-88-3881
IBAN: GB64NDCS47020318571282

Client:
Martinez-English
12196 Barrett Plain Suite 873
North Victorborough, NY 32585

Tax Id: 965-87-8819

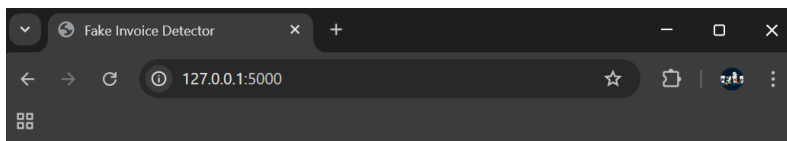
ITEMS

No.	Description	Qty	UM	Net price	Net worth	VAT [%]	Gross worth
1.	4"x2" Black Center Top Table Italian Marquetry Gemstone Art Columba's Day Gift	2,00	each	3 688,00	7 376,00	10%	8 113,60
2.	12" Marble Side Coffee Table Top Multi Stone Floral Inlay Occasional Decor W260	4,00	each	259,77	1 039,08	10%	1 142,99

SUMMARY

	VAT [%]	Net worth	VAT	Gross worth
	10%	8 415,08	841,51	9 256,59
Total		\$ 8 415,08	\$ 841,51	\$ 9 256,59

Output:



Fake Invoice Detector

Choose file No file chosen

Analyze Invoice

Extracted Fields:

- **Invoice No:** 51295021
- **Date of Issue:** 01/05/2017
- **Seller Tax ID:** 922-88-3881
- **Client Tax ID:** 965-87-8819
- **IBAN:** GB64NDCS47020318571282

Original Invoice