

Smart Invoice Automation Project Report

Using n8n, OCR Space API, AI Extractor & Google Sheets

Page 1: Executive Summary

Project Title: Smart Invoice Data Extraction using n8n + AI

Objective:

To build a fully automated solution that extracts structured invoice data (Invoice No, Date, Items, Amounts, etc.) from both **digital** and **scanned PDFs** and stores it in **Google Sheets** — without any manual steps.

Key Highlights:

- Handles scanned PDFs using **OCR Space API**
 - Parses digital PDFs using **native PDF extract**
 - Structured JSON output via **AI Information Extractor**
 - Scalable workflow using **n8n**
-





Page 2: Problem Statement & Use Cases

2.1 The Challenge

Manual invoice processing is:

- Time-consuming ⌚
- Error-prone ❌
- Non-scalable for large volumes 📉
- Difficult when invoices are **scanned images** (no text layer)

2.2 Real-World Use Cases:

-  **Finance teams** – extracting data from hundreds of monthly vendor bills
 -  **Startups** – automating invoice reconciliation
 -  **Healthcare or Retail** – batch processing of scanned supplier invoices
 -  **Freelancers** – logging client invoices directly to sheets for reporting
-

Page 3: Sample Invoices (Input)

3.1 Invoice Type 1 – Scanned (Text Not Extractable)

- **File:** INV-1001.pdf & INV-1002.pdf
- **Format:** Scanned image-only PDF
- **Challenge:** Cannot extract text with native PDF extract

3.2 Invoice Type 2 – Digital PDF (Text Extractable)

- **File:** invoice-0-4.pdf
 - **Extracted Invoice Info:**
 - **Invoice #:** BPXINV-00550
 - **Date:** 23.05.2021
 - **Customer:** Roger Bigot
 - **Items:** 30+ items listed with quantity, unit price & total
 - **Total Due:** ₹6610.95
-

Page 4: Workflow Architecture (n8n Flow)

Step 1: Manual Trigger



Step 2: Google Drive – Download File



Step 3: Extract Text from PDF



Step 4: IF (Is text empty?)



OCR API

Info Extractor



-----> Google Sheets Append

Visual Reference: Flowchart diagram (see image asset 2025-07-26T09-25-02.874Z.png)

Page 5: Workflow Step-by-Step with Parameters

Step 1 – Manual Trigger

- **Purpose:** Start the workflow manually or later schedule it.
-

Step 2 – Google Drive Download

- **Operation:** Download
 - **File Name:** INV-1001.pdf
 - **Mime Type:** application/pdf
 - **Binary Output Field:** data
-

Step 3 – Extract from PDF

- **Operation:** Extract From PDF
- **Input Binary Field:** data
- **Output:**
 - text, numpages, info, PDFFormatVersion

If text is missing or blank → go to OCR.

Step 4 – IF Node

- **Condition:**

{{ \$json.text }} is empty

- **Branches:**
 - True → Step 5A (OCR)
 - False → Step 5B (AI Extractor)
-

Page 6: OCR Space Integration (Step 5A)

Node: HTTP Request

- **Method:** POST
- **URL:** https://api.ocr.space/parse/image
- **Headers:**
 - apikey: K87032385688957
- **Body:**

- base64Image:

data:{{ \$('Download file').item.binary.data.mimeType }};base64,{{ \$('Download file').item.binary.data.data }}

- language: eng
- isOverlayRequired: false

Output:

json

CopyEdit

```
{
  "ParsedText": "Invoice INV-1003\nBill To: Orbit Electronics\n...",
  "invoice_date": "2025-06-07",
  "items": [...]
}
```



Page 7: AI Information Extractor (Step 5B)

Input Prompt to AI:

“Extract key fields like invoice number, customer name, items, quantity, total, and output it in JSON format.”

AI Output:

```
{
  "invoice_number": "INV-1003",
  "invoice_date": "2025-06-07",
  "bill_to": {
    "name": "Orbit",
    "company": "Electronics"
  },
  "items": [
    {
      "description": "Gadget B",
      "quantity": 8,
      "unit_price": 34.5,
```

```
"total": 276
}
],
"total_due": 693.73
}
```

Page 8: Google Sheets Integration (Step 6)

Node: Google Sheets – Append Row






- **Document:** Invoice_data_Extraction
- **Sheet:** Sheet1
- **Mapped Columns:**
 - Invoice Number → {{ \$json.output.invoice_number }}
 - Description → {{ \$json.output.items[0].description }}
 - Quantity → {{ \$json.output.items[0].quantity }}
 - Unit Price → {{ \$json.output.items[0].unit_price }}
 - Total Price → {{ \$json.output.items[0].total }}

Sample Output:

Invoice No	Date	Customer	Description	Qty	Price	Total
INV-1003	2025-06-07	Orbit	Gadget B	8	34.5	276

Page 9: Benefits of the Solution

Functional Benefits:

-  No manual copy-paste
-  Works with both digital & scanned PDFs
-  Intelligent field extraction using AI
-  Real-time Google Sheets sync
-  Reusable across invoice formats and clients

Business Impact:

- Saves **2–4 hours** daily for finance/admin staff
- Reduces **entry errors by 95%**

- Helps in **faster reconciliation and reporting**
-

Page 10: Conclusion & Future Scope

Conclusion:

This project demonstrates how **no-code tools** like n8n, when combined with **AI and OCR APIs**, can **fully automate repetitive business processes**.

It replaces a highly manual task — invoice logging — with a smart system that:

- Detects file type
- Applies appropriate extraction technique
- Saves final output instantly

Future Scope:

- Auto-fetch from email (e.g., Gmail with attachments)
- Schedule daily/weekly batch jobs
- Log failed extractions to Slack/email
- Store output in MySQL/BigQuery for analytics
- Generate reports or dashboards (Power BI, Looker Studio)