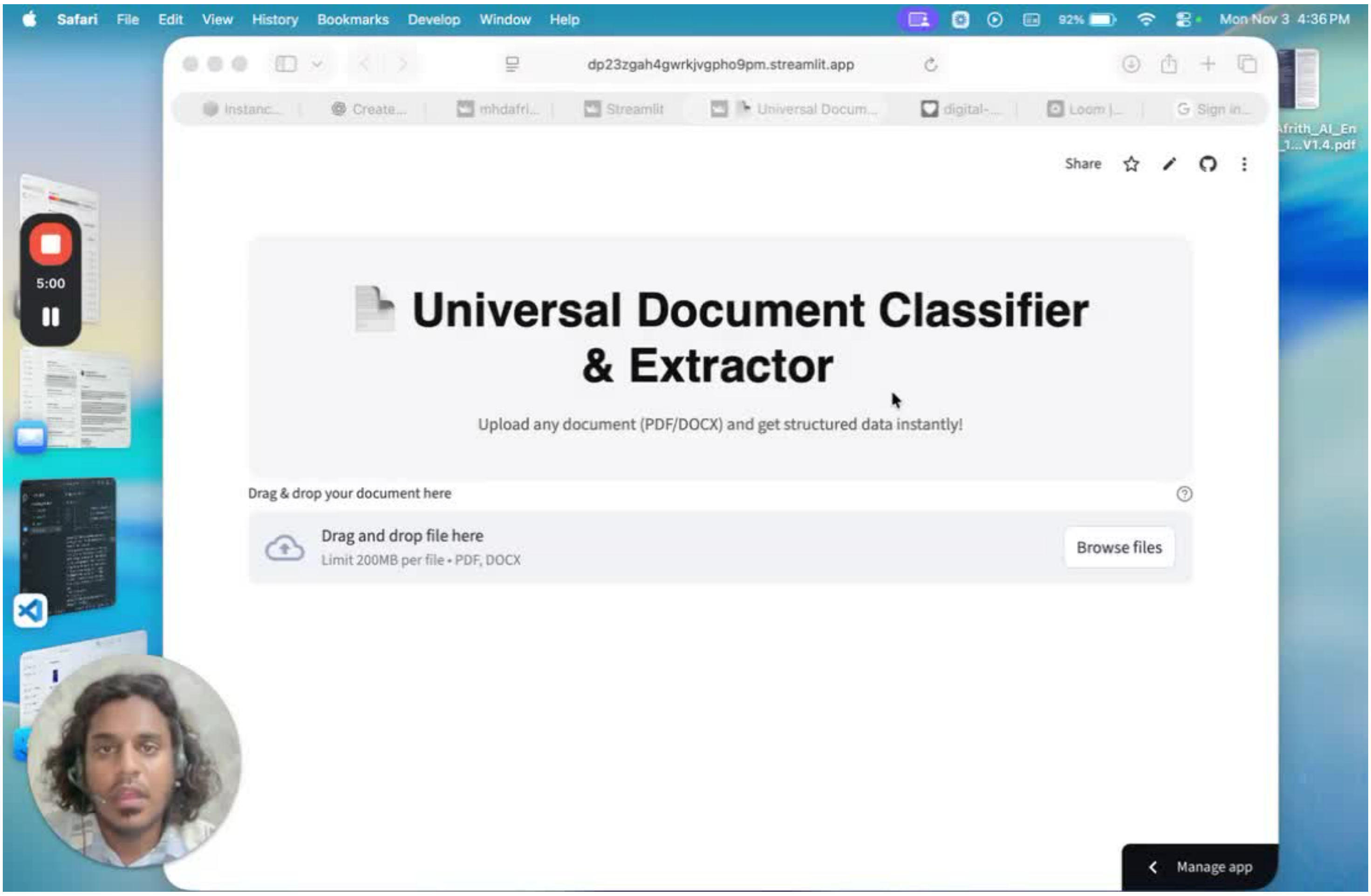
Universal Document Classifier and Extractor SOP

Objective

This SOP outlines the steps to use the Universal Document Classifier and Extractor application for automated document classification and information extraction.

Key Steps

1. Introduction to the Application 0:01



Universal Document Classifier & Extractor

Upload any document (PDF/DOCX) and get structured data instantly!

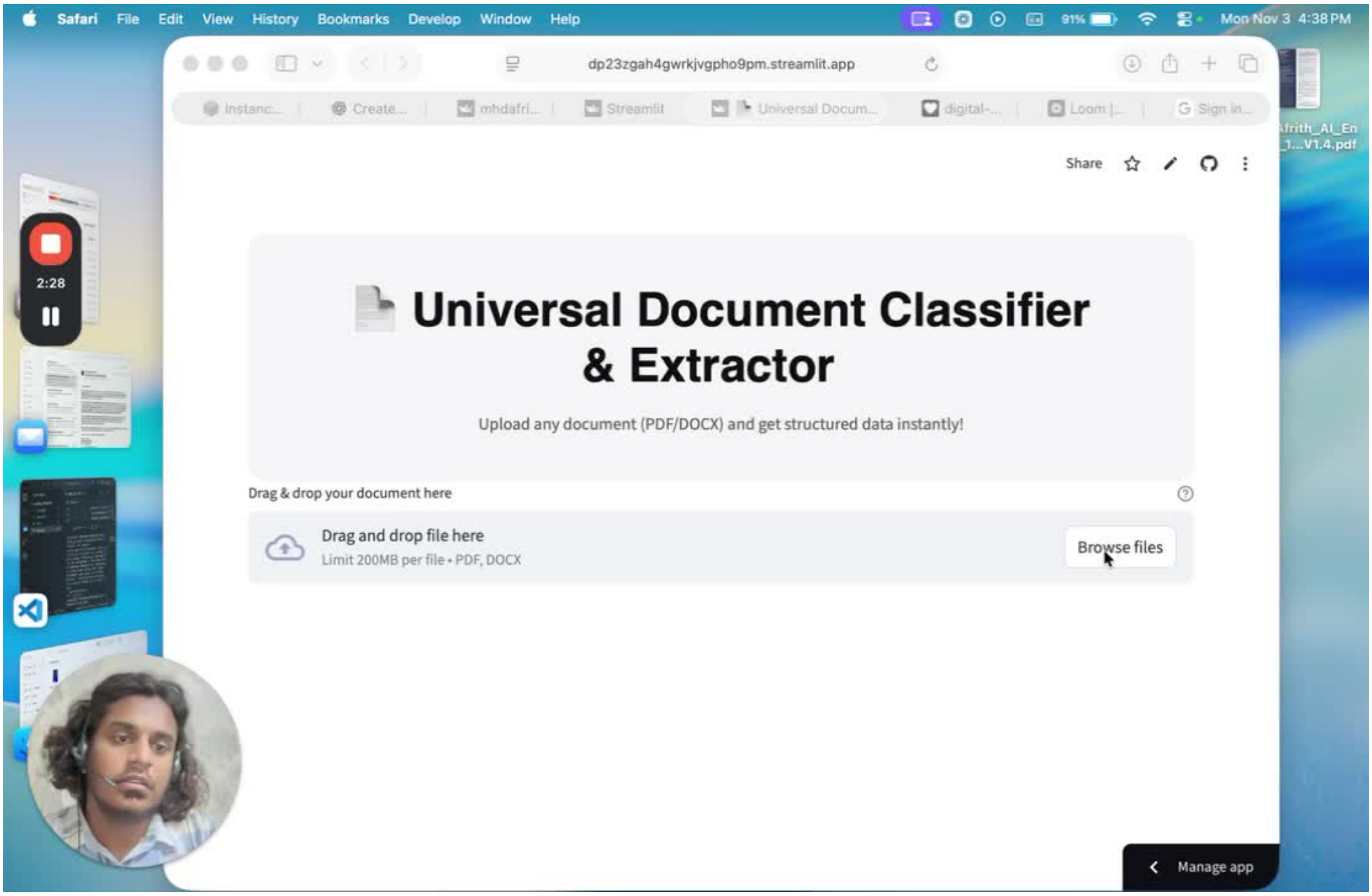
Drag & drop your document here

Drag and drop file here

Limit 200MB per file • PDF, DOCX

- The Universal Document Classifier and Extractor is an AI-powered application designed to classify and extract key information from various document types, including:
    - Invoices
    - Resumes
    - Bank statements
    - Medical reports
- The goal is to reduce manual document processing and automate information extraction.

2. Uploading a Document 2:03



Safari File Edit View History Bookmarks Develop Window Help
Mon Nov 3 4:38 PM
dp23zgah4gwrkjvgpho9pm.streamlit.app
Instanc... Create... mhdafri... Streamlit Universal Docum... digital-... Loom |... Sign in...
Share
2:28
Manage app
Browse files

Universal Document Classifier & Extractor

Upload any document (PDF/DOCX) and get structured data instantly!

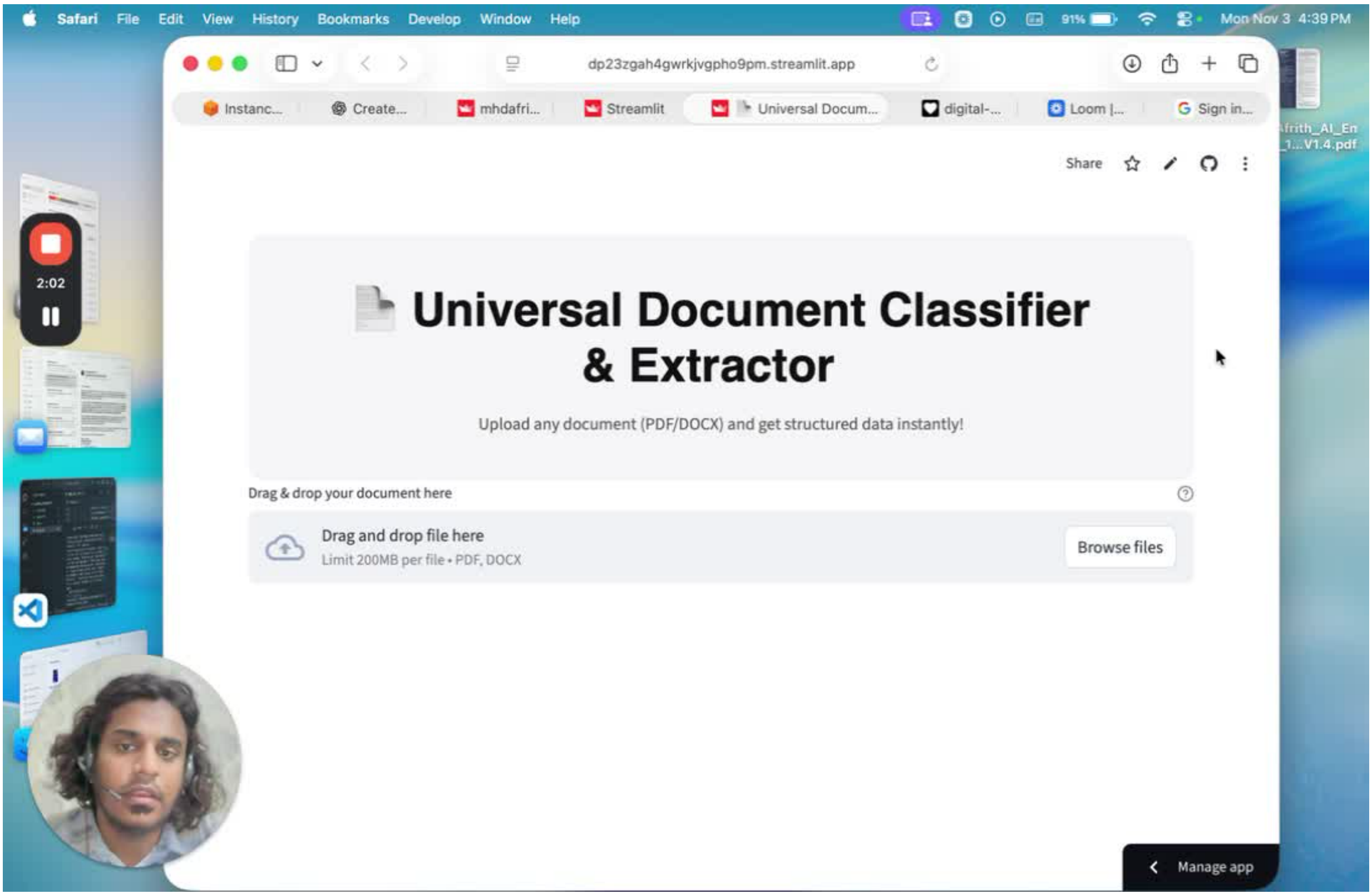
Drag & drop your document here

Drag and drop file here

Limit 200MB per file • PDF, DOCX

- To begin using the application, follow these steps:
  - Navigate to the upload section of the application.
  - Click on the upload button and select the document you wish to classify (e.g., a resume).

3. Document Classification Process 2:28



Safari File Edit View History Bookmarks Develop Window Help
Mon Nov 3 4:39 PM
dp23zgah4gwrkjvgpho9pm.streamlit.app
Instanc... Create... mhdafri... Streamlit Universal Docum... digital-... Loom |... Sign in...
Share
2:02
Manage app

Universal Document Classifier & Extractor

Upload any document (PDF/DOCX) and get structured data instantly!

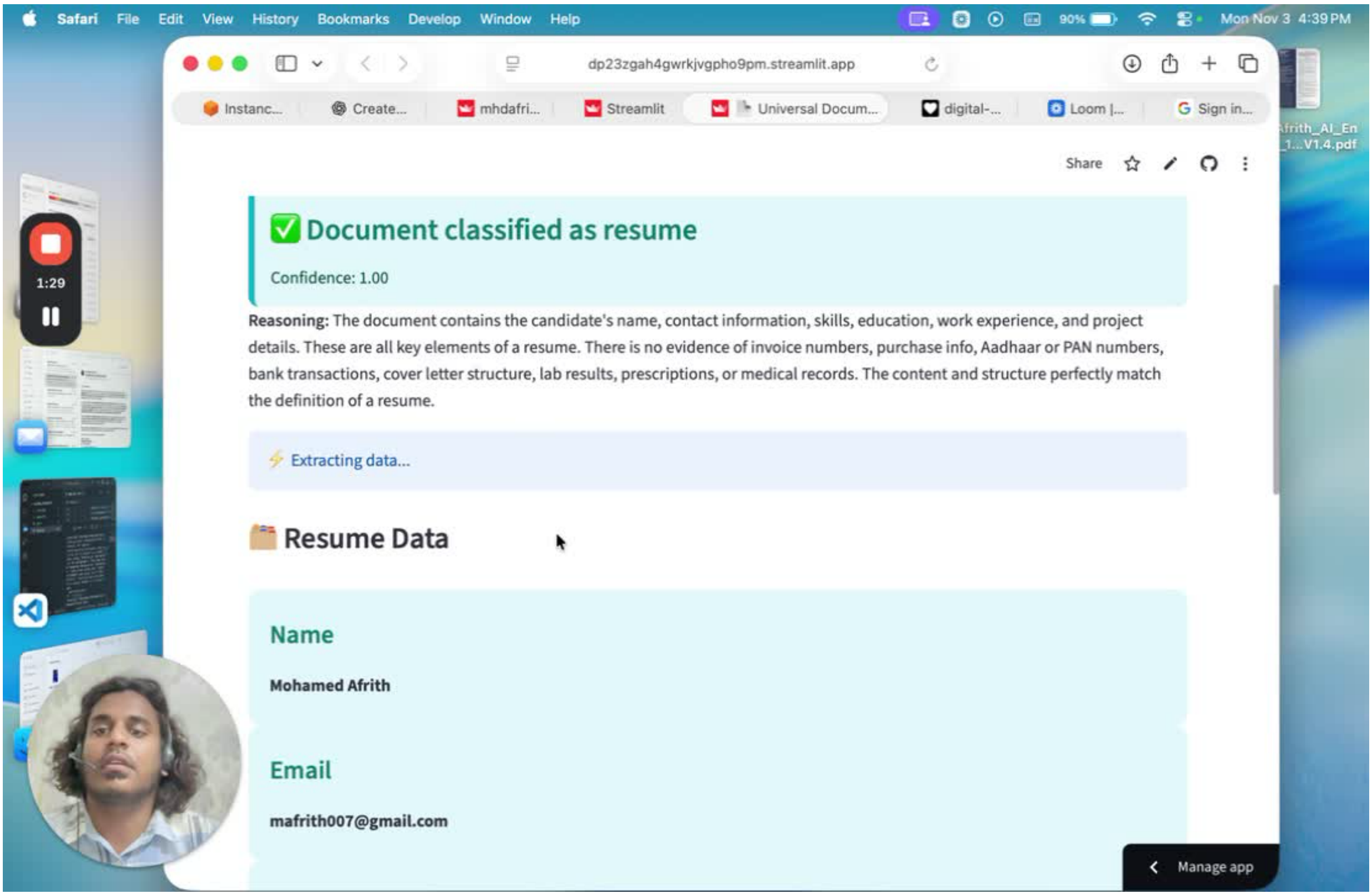
Drag & drop your document here

Drag and drop file here
Limit 200MB per file • PDF, DOCX

Browse files

- Once the document is uploaded, the application will:
  - Identify the type of document (e.g., resume).
  - Provide reasoning for the classification, indicating how confident the system is in its classification.

4. Information Extraction 3:01



Safari File Edit View History Bookmarks Develop Window Help
Mon Nov 3 4:39 PM
dp23zgah4gwrkjvgpho9pm.streamlit.app
Instanc... Create... mhdafri... Streamlit Universal Docum... digital-... Loom |... Sign in...
Share
1:29
Manage app

✅ Document classified as resume

Confidence: 1.00

Reasoning: The document contains the candidate's name, contact information, skills, education, work experience, and project details. These are all key elements of a resume. There is no evidence of invoice numbers, purchase info, Aadhaar or PAN numbers, bank transactions, cover letter structure, lab results, prescriptions, or medical records. The content and structure perfectly match the definition of a resume.

⚡ Extracting data...

Resume Data

Name

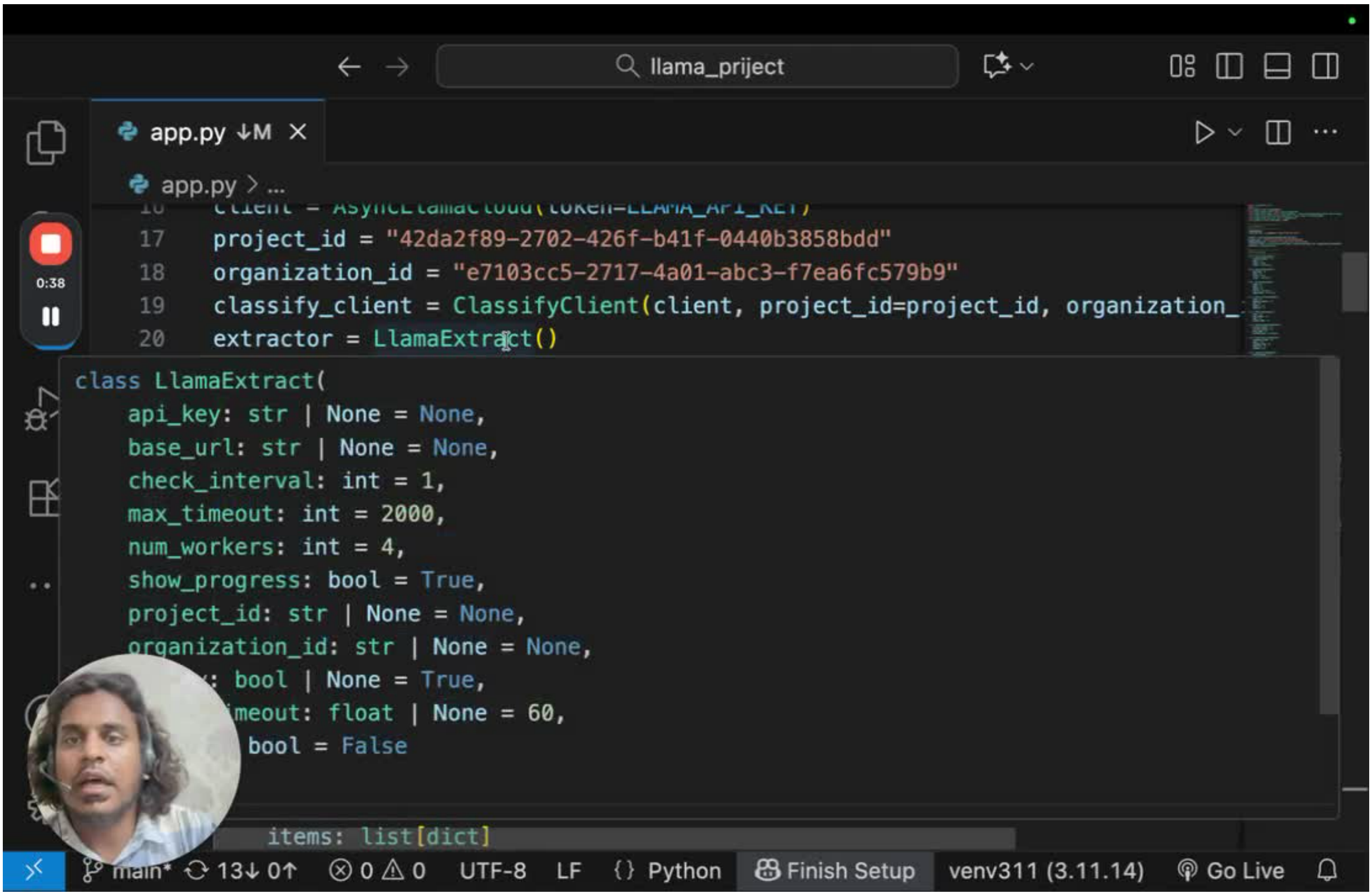
Mohamed Afrith

Email

mafrith007@gmail.com

- After classification, the application will:
  - Activate the extraction agent to retrieve key values from the document.
  - Present the extracted information in a structured format.

5. Code Overview 3:44



```
client = AsyncLlamaCloud(token=LLAMA_API_KEY)
project_id = "42da2f89-2702-426f-b41f-0440b3858bdd"
organization_id = "e7103cc5-2717-4a01-abc3-f7ea6fc579b9"
classify_client = ClassifyClient(client, project_id=project_id, organization_
extractor = LlamaExtract()
```

```
class LlamaExtract(
    api_key: str | None = None,
    base_url: str | None = None,
    check_interval: int = 1,
    max_timeout: int = 2000,
    num_workers: int = 4,
    show_progress: bool = True,
    project_id: str | None = None,
    organization_id: str | None = None,
        : bool | None = True,
       imeout: float | None = 60,
         bool = False

    items: list[dict]
```

- The application uses specific libraries and APIs for functionality:
    - Import necessary libraries.
    - Load the Llama Cloud API for document classification and extraction.
    - Define schemas for different document classes (e.g., invoices, resumes).

Cautionary Notes

- Ensure that the documents uploaded do not contain sensitive personal information unless necessary for processing.
- Verify the accuracy of the extracted information, as automated systems may not always be 100% accurate.

Tips for Efficiency

- Regularly update the document classification schemas to include new document types as needed.
- Train the system with diverse examples to improve classification accuracy.

Link to Loom

https://loom.com/share/67edd2870660465895dce4fde7726eb3