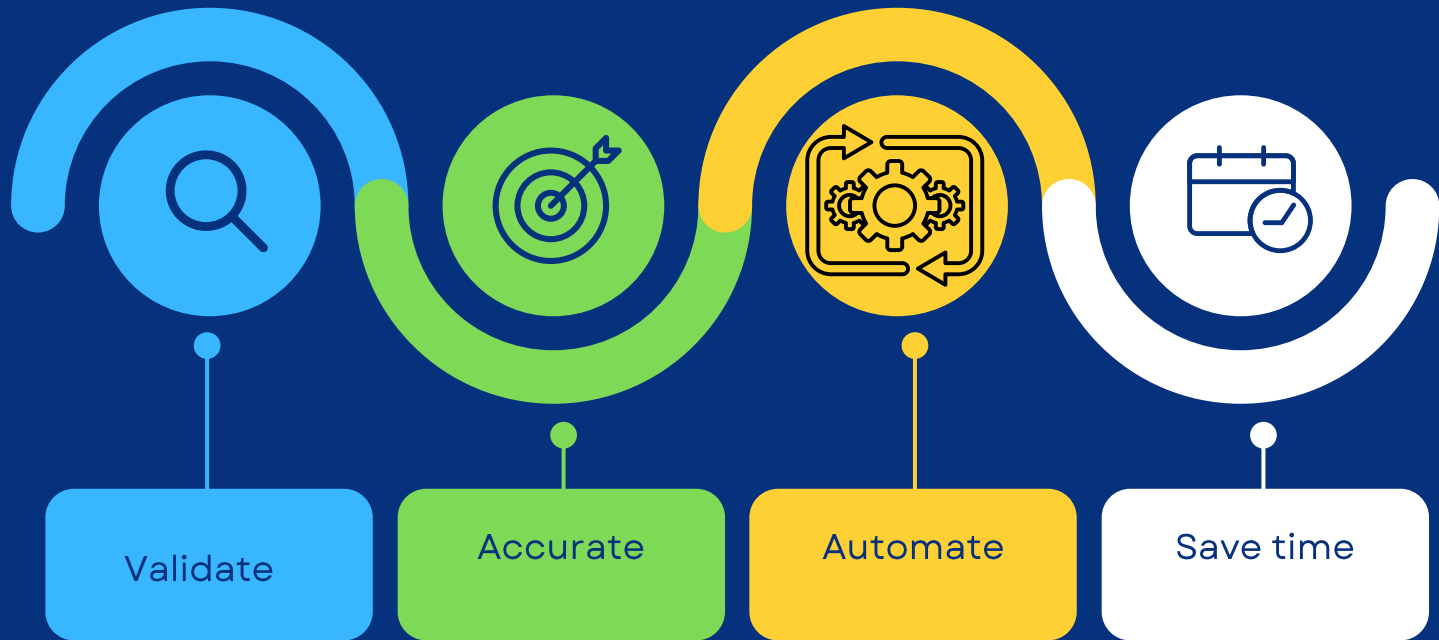




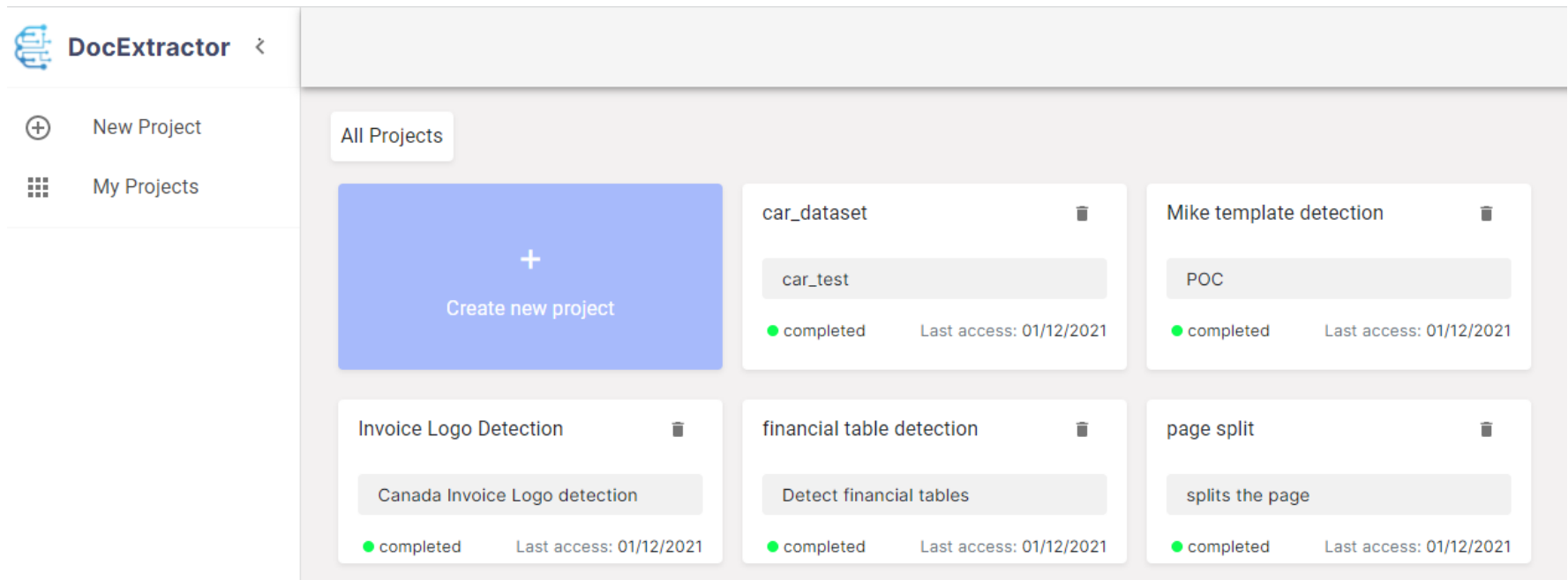
Docextractor

Intelligent Document Processing

A Step by Step Guide



Create Project

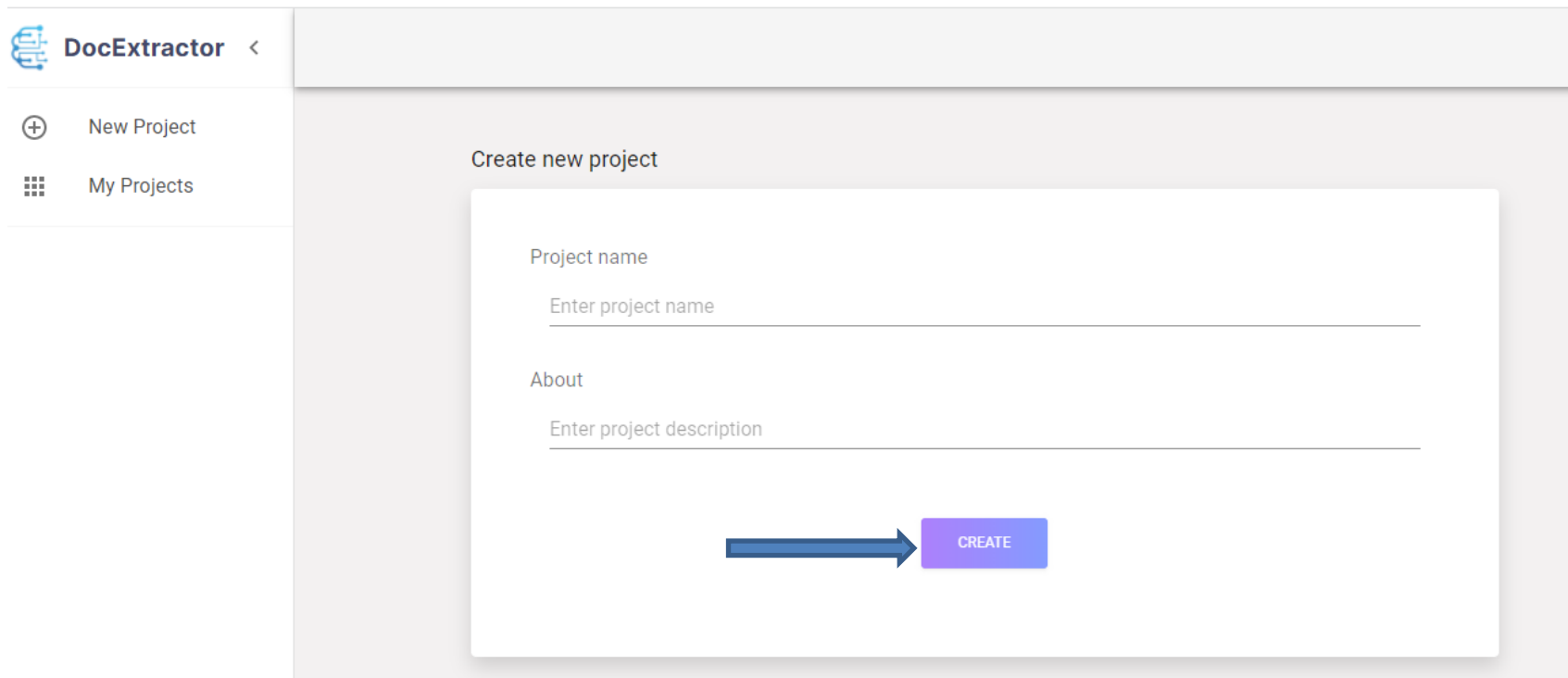


The screenshot displays the DocExtractor application interface. On the left, a sidebar contains the DocExtractor logo and navigation options: 'New Project' (with a plus icon) and 'My Projects' (with a grid icon). The main content area is titled 'All Projects' and features a large blue button with a plus sign and the text 'Create new project'. Below this, there are six project cards arranged in a 2x3 grid. Each card has a title, a trash icon, a sample text box, a status indicator (a green dot followed by 'completed'), and a 'Last access' timestamp of '01/12/2021'. The projects are: 'car_dataset' (sample: 'car_test'), 'Mike template detection' (sample: 'POC'), 'Invoice Logo Detection' (sample: 'Canada Invoice Logo detection'), 'financial table detection' (sample: 'Detect financial tables'), and 'page split' (sample: 'splits the page').

Project Name	Sample Text	Status	Last Access
car_dataset	car_test	completed	01/12/2021
Mike template detection	POC	completed	01/12/2021
Invoice Logo Detection	Canada Invoice Logo detection	completed	01/12/2021
financial table detection	Detect financial tables	completed	01/12/2021
page split	splits the page	completed	01/12/2021

Click on “Create New project ” to create a project

Project Creation



The screenshot shows the DocExtractor application interface. On the left is a sidebar with the DocExtractor logo and a back arrow, and two menu items: 'New Project' with a plus icon and 'My Projects' with a grid icon. The main content area is titled 'Create new project' and contains a form with two text input fields. The first field is labeled 'Project name' with a placeholder 'Enter project name'. The second field is labeled 'About' with a placeholder 'Enter project description'. Below the form is a blue arrow pointing to a purple 'CREATE' button.

After providing Project Name and About click the the Create button

Create a New Model

The screenshot shows the DocExtractor web interface. On the left is a sidebar with the DocExtractor logo and navigation options: 'New Project' and 'My Projects'. The main content area is titled 'Demo-Invoice Processing'. A blue arrow points from the text to a purple 'NEW MODEL' button. To its right is another purple button labeled 'UPLOAD PRE-TRAINED MODEL'. Below these buttons, the text 'About Invoice Processing' is visible. At the bottom, a white box labeled 'Models' contains a red message: 'No model found'. A green notification box at the top right indicates 'Project created successfully'.

Create a New Model to process your document

Model Creation and Image Upload

DocExtractor <

⊕ New Project

☰ My Projects

🏠 / Create Model

Create a new model

Model name

Demo-Invoice processing

Labels

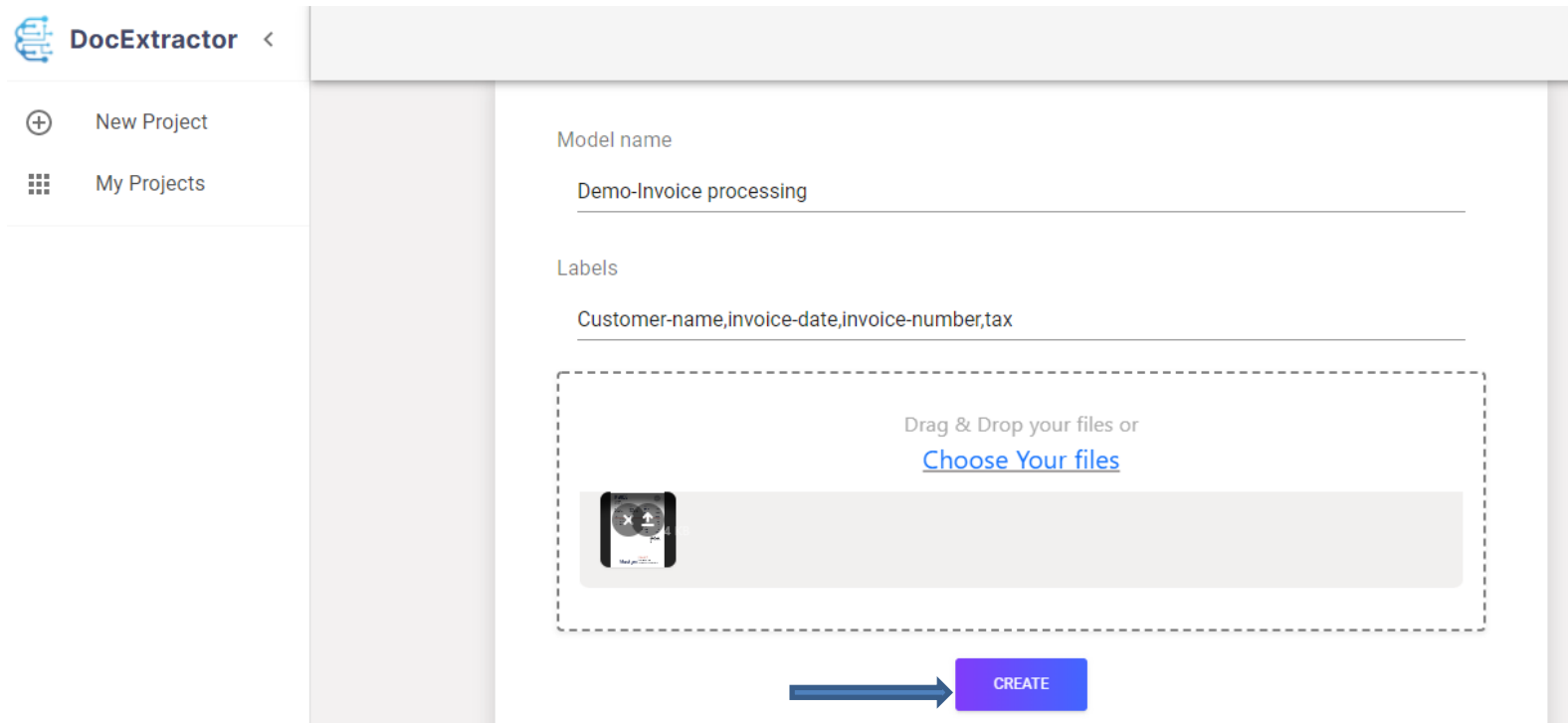
Customer-name,invoice-date,invoice-number,tax|

Drag & Drop your files or [Choose Your files](#)

Provide Model Name and Labels.

Label denotes the which data you want to extract from your document .

Process for Annotation/Labeling



The screenshot shows the DocExtractor web interface. On the left is a sidebar with the DocExtractor logo and navigation options: 'New Project' (with a plus icon) and 'My Projects' (with a grid icon). The main content area contains a form with the following fields:

- Model name:** A text input field containing 'Demo-Invoice processing'.
- Labels:** A text input field containing 'Customer-name,invoice-date,invoice-number,tax'.
- File Upload:** A dashed box containing the text 'Drag & Drop your files or [Choose Your files](#)'. Below this text is a grey rectangular area with a file upload icon (a document with a plus sign).

At the bottom right of the form, there is a blue button labeled 'CREATE'. A blue arrow points from the file upload area towards the 'CREATE' button.

Click Create button after images are added

Labelling of the image

The screenshot shows the DocExtractor web application interface. On the left, there is a sidebar with 'New Project' and 'My Projects' options. The main area displays an invoice image titled 'invoice-template-us-neat-750px.png'. A red bounding box is drawn over the 'BILL TO' section. A dropdown menu is open over this box, showing a 'Classification' field and a list of labels: 'Customer-name', 'invoice-date', 'invoice-number', and 'tax'. The 'Customer-name' label is currently selected. The invoice content includes the company name 'East Repair Inc.', address, and a table with columns 'QTY', 'DESCRIPTION', 'UNIT PRICE', and 'AMOUNT'. The table contains one row: '1 Front and rear brake cables' with a unit price of 100.00 and an amount of 100.00. The interface also features navigation buttons like 'PREV', 'NEXT', 'SETTINGS', 'FULLSCREEN', and 'SAVE', along with a 'DELETE IMAGE' button and a 'TRAIN' button.

Create Bounding box over the region which needs to be extracted, The labels you have provided in the last screen will automatically come in the drop down .

Train of the Model

The screenshot shows the DocExtractor application interface. On the left is a sidebar with 'New Project' and 'My Projects'. The main area displays an invoice template titled 'invoice-template-us-neat-750px.png'. A progress bar at the top indicates '1/1 done' and '100%'. A red 'DELETE IMAGE' button and a blue 'TRAIN' button are visible. A blue arrow points to the 'TRAIN' button. The invoice content includes:

INVOICE

East Repair Inc.
1912 Harvest Lane
New York, NY 12210

BILL TO
John Smith
2 Court Square
New York, NY 12210

SHIP TO
Customer-name
John Smith
3787 Pineview Drive
Cambridge, MA 12210

LOGO

invoice-number: US-001
invoice-date: 11/02/2019
P.O.#: 2312/2019
DUE DATE: 26/02/2019

QTY	DESCRIPTION	UNIT PRICE	AMOUNT
-----	-------------	------------	--------

Click the Train button after the Labelling of the region is done

Model after Training

The screenshot displays the DocExtractor application interface. On the left, a sidebar contains the DocExtractor logo and navigation options: 'New Project' and 'My Projects'. The main content area shows the breadcrumb path 'Demo-Invoice Processing / Demo-Invoice processing'. Below this, a blue button labeled 'Demo-Invoice processing' is visible. Underneath, the 'Labels:' section lists four categories: 'Customer-name', 'invoice-date', 'invoice-number', and 'tax'. At the bottom, the 'Training Status' section shows 'Model is untrained', with a blue arrow pointing to this text.

Model Training status will show here

Model Status and Metrics

The screenshot displays the DocExtractor web interface. On the left is a sidebar with the DocExtractor logo and navigation options: 'New Project' and 'My Projects'. The main content area shows the breadcrumb 'Invoice Extraction V1 EU / invoice model v2'. Below this, a blue button labeled 'invoice model v2' is visible. Underneath, the 'Labels:' section lists five categories: 'customer_name', 'customer_address', 'invoice_no', 'invoice_date', and 'total'. The 'Training Status' section indicates 'Model training is completed' and features two blue buttons: 'VIEW MODEL METRICS' and 'TEST MODEL'.

After the model is trained you click “VIEW MODEL METRICES” to check the accuracy of the model

API Generation

The screenshot displays the DocExtractor web interface. On the left, a sidebar contains the DocExtractor logo, a 'New Project' button with a plus icon, and a 'My Projects' button with a grid icon. The main content area is divided into two sections. The top section, 'Training Status', shows 'Model training is completed' and includes two blue buttons: 'VIEW MODEL METRICS' and 'TEST MODEL'. The bottom section, 'API Details', shows a 'POST' method. The 'Request URL' is 'https://demo.docextractor.com/api/ocr/extract/628f2a84bd227b7d97b32bc5', with a copy icon and a blue arrow pointing to it. The 'Request Body' is 'files : [list of image files]'.

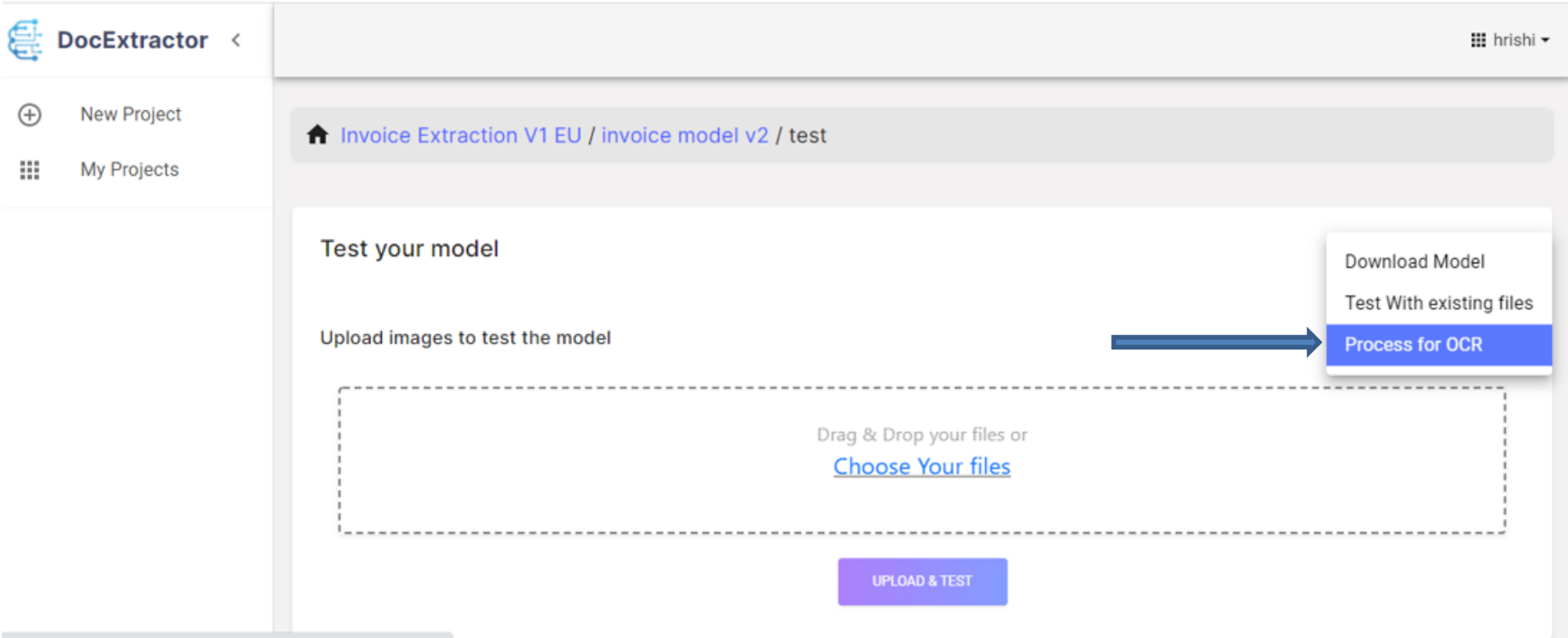
Automatic API generation , the api generated can be directly integrated with your system for processing of your document

Testing the Model

The screenshot displays the DocExtractor web interface. On the left, a sidebar contains the DocExtractor logo and navigation options: 'New Project' and 'My Projects'. The main content area shows the breadcrumb 'Invoice Extraction V1 EU / invoice model v2'. Below this, a blue button labeled 'invoice model v2' is visible. Underneath, a 'Labels:' section lists five categories: 'customer_name', 'customer_address', 'invoice_no', 'invoice_date', and 'total'. The 'Training Status' section indicates 'Model training is completed' and features two buttons: 'VIEW MODEL METRICS' and 'TEST MODEL'. A blue arrow points to the 'TEST MODEL' button, indicating the next step in the process.

Click Test Model to test the accuracy of extraction of data

OCR Output



Upload image and test the output by clicking “Process for OCR”

www.docextractor.com

email :- ananya@docextractor.com