

# IBM Automation Document Processing

Business Analyst Training

Livret d'exercices

## Table des matières

Introduction.....	3
Serveurs IBM Cloud Pak for Business Automation .....	4
Exercice 1 - Connexion à Cloud Pak For Business Automation.....	5
Exercice 2 – Création d'un projet.....	8
Exercice 3 – Ajout d'une nouvelle classification .....	13
Exercice 4 – Entraînement du modèle de classification .....	22
Exercice 5 – Validation du modèle de classification .....	27
Exercice 6 – Modèle d'extraction .....	30

# Introduction

Bienvenue dans la formation IBM Document Processing pour Business Analysts.

Ces exercices vous présenteront comment vous familiariser à l'environnement IBM Cloud Pak, configurer le traitement automatisé des documents, la classification, et l'extraction.

Cette formation ne couvrira pas toutes les fonctionnalités disponibles en raison de contraintes de temps. Il est conçu comme un point d'entrée, une initiation.

## Serveurs IBM Cloud Pak for Business Automation

Participant	Environnement	Utilisateur	Mot de passe
Isabelle	<a href="https://cpd-cp4ba.apps.66d8a92d4596c2c4505a8245.ocp.techzone.ibm.com">https://cpd-cp4ba.apps.66d8a92d4596c2c4505a8245.ocp.techzone.ibm.com</a>	cpadmin	axgVt-tHyoG-WTZcF-gPPhd
Delphine	<a href="https://cpd-cp4ba.apps.66d8a92d4596c2c4505a8245.ocp.techzone.ibm.com">https://cpd-cp4ba.apps.66d8a92d4596c2c4505a8245.ocp.techzone.ibm.com</a>	cpadmin	axgVt-tHyoG-WTZcF-gPPhd
Wendy	<a href="https://cpd-cp4ba.apps.66d9c918d361e1cd7ea1d83e.ocp.techzone.ibm.com">https://cpd-cp4ba.apps.66d9c918d361e1cd7ea1d83e.ocp.techzone.ibm.com</a>	cpadmin	leUjC-IBpSE-pQfsM-lXbHr
Florence	<a href="https://cpd-cp4ba.apps.66d9c918d361e1cd7ea1d83e.ocp.techzone.ibm.com">https://cpd-cp4ba.apps.66d9c918d361e1cd7ea1d83e.ocp.techzone.ibm.com</a>	cpadmin	leUjC-IBpSE-pQfsM-lXbHr
Benoit	<a href="https://cpd-cp4ba.apps.66d7f605f6599fbacff450ef.ocp.techzone.ibm.com">https://cpd-cp4ba.apps.66d7f605f6599fbacff450ef.ocp.techzone.ibm.com</a>	cpadmin	cCLiz-wZYHF-xEPDI-SCvpD
Zakariya	<a href="https://cpd-cp4ba.apps.66d7f605f6599fbacff450ef.ocp.techzone.ibm.com">https://cpd-cp4ba.apps.66d7f605f6599fbacff450ef.ocp.techzone.ibm.com</a>	cpadmin	cCLiz-wZYHF-xEPDI-SCvpD
Luca	<a href="https://cpd-cp4ba.apps.66d9a1b8f6599fbacff453b5.ocp.techzone.ibm.com">https://cpd-cp4ba.apps.66d9a1b8f6599fbacff453b5.ocp.techzone.ibm.com</a>	cpadmin	PPDuN-RnFPc-TnarP-hyAYw
Michael	<a href="https://cpd-cp4ba.apps.66d9a1b8f6599fbacff453b5.ocp.techzone.ibm.com">https://cpd-cp4ba.apps.66d9a1b8f6599fbacff453b5.ocp.techzone.ibm.com</a>	cpadmin	PPDuN-RnFPc-TnarP-hyAYw
Tom	<a href="https://cpd-cp4ba.apps.66ddb81d361e1cd7ea1df0f.ocp.techzone.ibm.com">https://cpd-cp4ba.apps.66ddb81d361e1cd7ea1df0f.ocp.techzone.ibm.com</a>	cpadmin	OsdxZ-farJj-PyiKo-BmuuG
		cpadmin	OsdxZ-farJj-PyiKo-BmuuG

⚠ Ces serveurs sont hébergés dans un centre de données appartenant à IBM en Europe.  
Il est fortement déconseillé d'y sauvegarder des documents de production pour des raisons de sécurité.

# Exercice 1 - Connexion à Cloud Pak For Business Automation

## Objectifs :

Dans cet exercice, nous allons vous expliquer comment se connecter à IBM Cloud Pak For Business Automation (ICP4BA), la plateforme qui héberge la solution IBM Automation Document Processing (ADP).

## Prérequis:

- Avant de commencer, veuillez consulter le tableau « **Serveurs IBM Cloud Pak for Business Automation** », reprenant la liste des environnements disponibles, afin de déterminer celui qui vous a été attribué, votre nom d'utilisateur (**cpadmin**), et le mot de passe correspondant.

Dans le cas d'Isabelle, le serveur est :

<https://cpd-cp4ba.apps.66d8a92d4596c2c4505a8245.ocp.techzone.ibm.com>

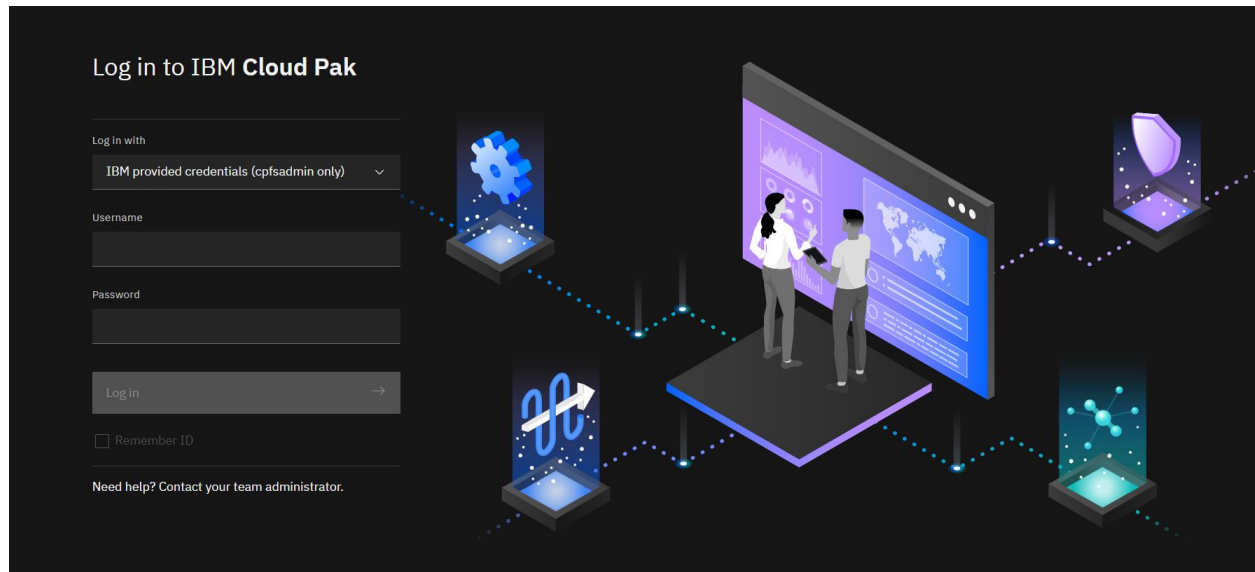
Et pour Benoit : <https://cpd-cp4ba.apps.66d7f605f6599fbacff450ef.ocp.techzone.ibm.com>

**Il est important que chacun utilise l'environnement qui lui a été assigné.**

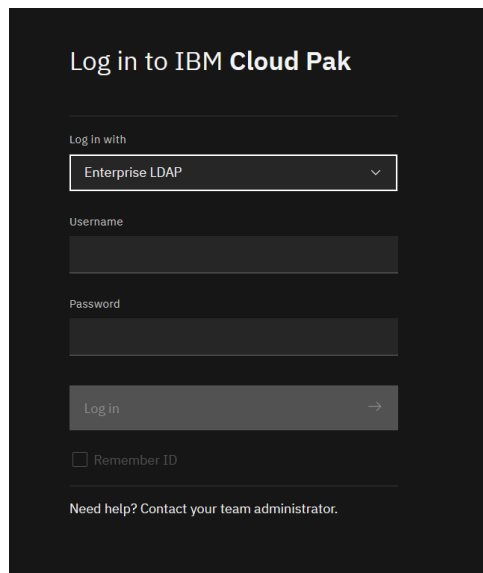
- Pour effectuer les exercices, il est fortement recommandé d'utiliser Google Chrome, ou Microsoft Edge.

\_1. Pour accéder à votre environnement ICP4BA, veuillez démarrer **Google Chrome** ou **Microsoft Edge** et veuillez ouvrir l'adresse du serveur ICP4BA qui vous été attribué.

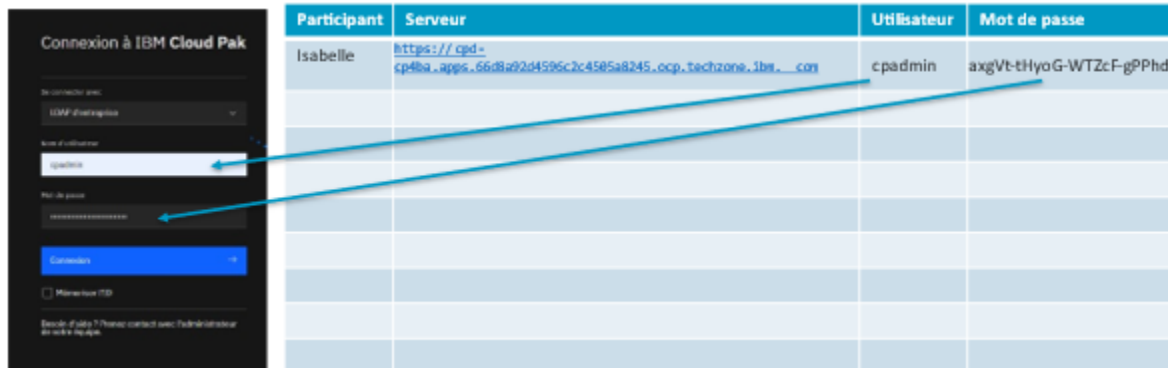
La page suivante devrait s'afficher :



\_2. Veuillez sélectionner Enterprise LDAP ou LDAP d'entreprise comme mode d'authentification.



\_3. Ensuite, à l'aide du tableau reprenant la liste des Serveurs IBM Cloud Pak for Business Automation, veuillez entrer votre utilisateur (**cpadmin**), et le mot de passe dans le formulaire de connexion.

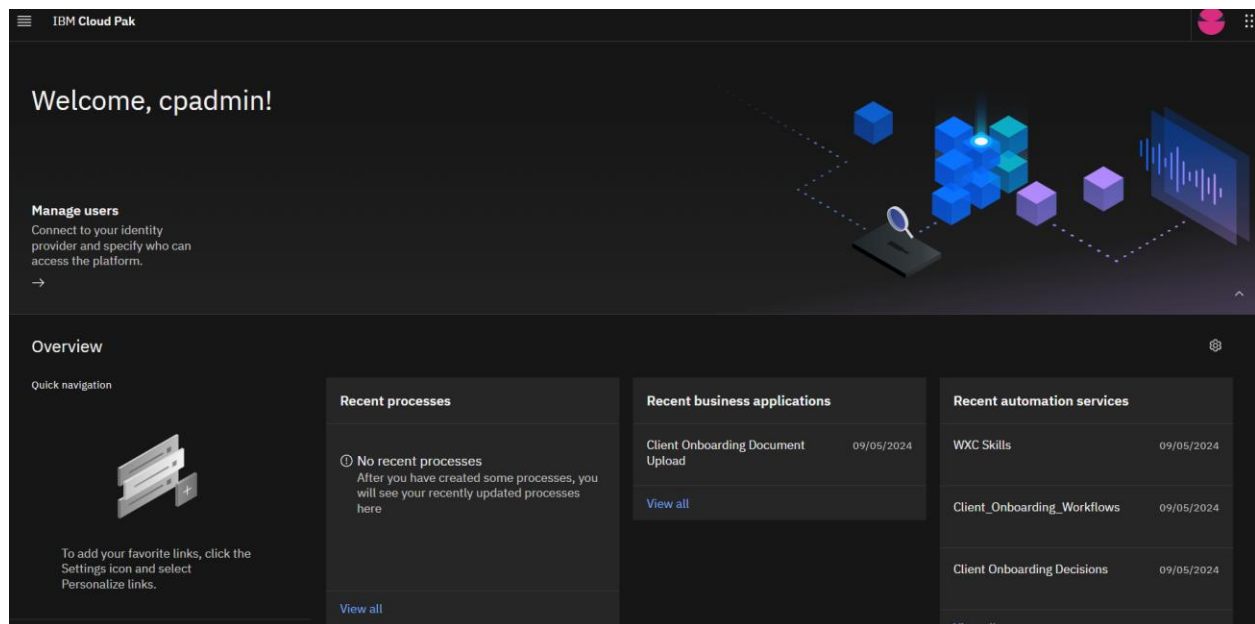


Participant	Serveur	Utilisateur	Mot de passe
Isabelle	<a href="https://cpd-cp4ba.apps.66d8a2d4596c2c4585a8245.ocp.techzone.ibm.com">https://cpd-cp4ba.apps.66d8a2d4596c2c4585a8245.ocp.techzone.ibm.com</a>	cpadmin	axgVt-tHyoG-WTZcF-gPPhd

\_4. Après avoir entré vos informations, cliquez sur le bouton « **Login** » ou « **Connexion** » en français.



\_5. Si vos identifiants ont été entrés correctement, l'écran d'accueil devrait s'afficher



IBM Cloud Pak

Welcome, cpadmin!

**Manage users**  
Connect to your identity provider and specify who can access the platform.  
→

**Overview**

Quick navigation

To add your favorite links, click the Settings icon and select Personalize links.

**Recent processes**

ⓘ No recent processes  
After you have created some processes, you will see your recently updated processes here  
[View all](#)

**Recent business applications**

Client Onboarding Document Upload 09/05/2024  
[View all](#)

**Recent automation services**

WXC Skills 09/05/2024  
Client\_Onboarding\_Workflows 09/05/2024  
Client Onboarding Decisions 09/05/2024  
[View all](#)

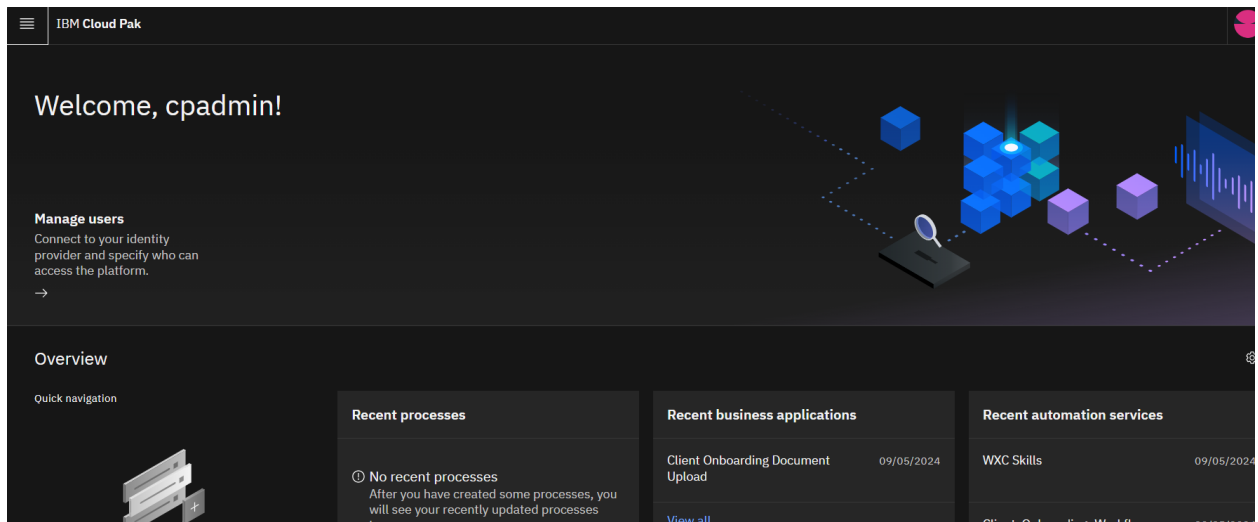
Félicitations, vous voilà connecté à votre environnement.


## Exercice 2 – Création d'un projet

### Objectifs :

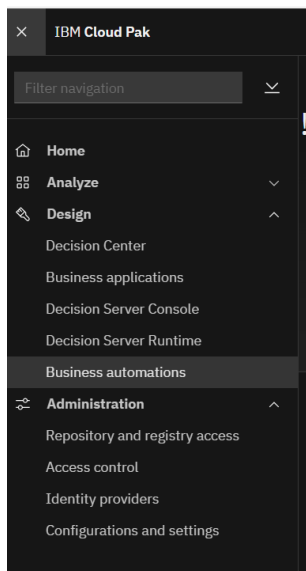
Dans cet exercice, vous allez apprendre à créer votre premier projet ADP, à l'aide de l'outil de conception, IBM Business Automation Studio.

\_1. Après authentification, et depuis la page d'accueil



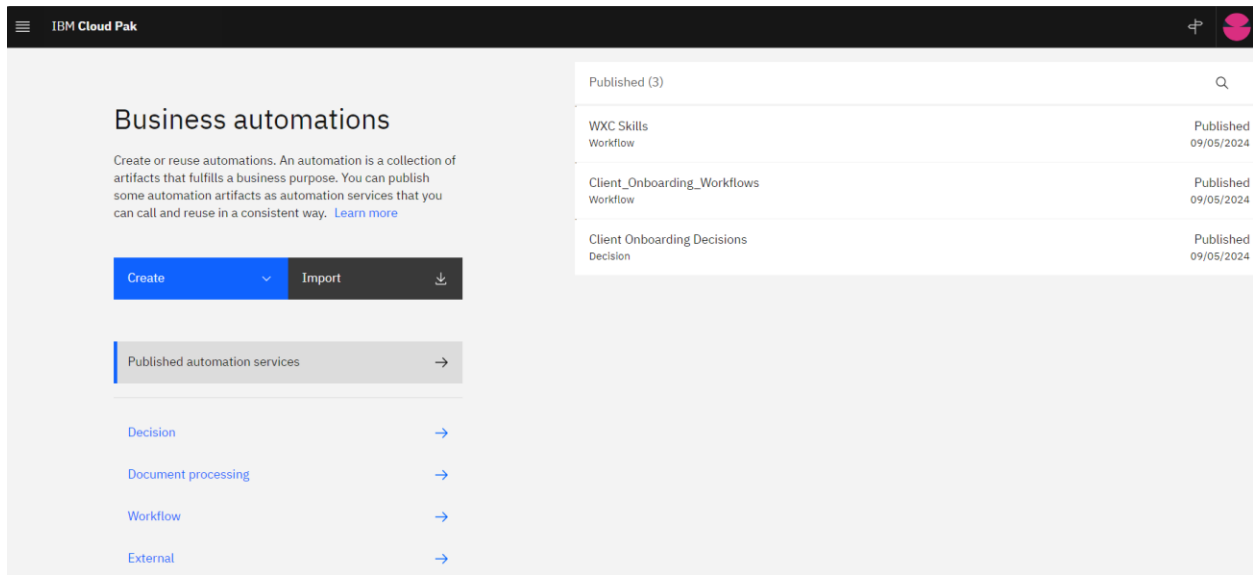
\_2. Cliquez sur l'hamburger  qui permet d'accéder aux différentes applications.

\_3. Et veuillez sélectionner « Business Automation » depuis le menu,





Vous arrivez dans l'écran d'accueil de l'outil de conception « Business Automation Studio »



IBM Cloud Pak

## Business automations

Create or reuse automations. An automation is a collection of artifacts that fulfills a business purpose. You can publish some automation artifacts as automation services that you can call and reuse in a consistent way. [Learn more](#)

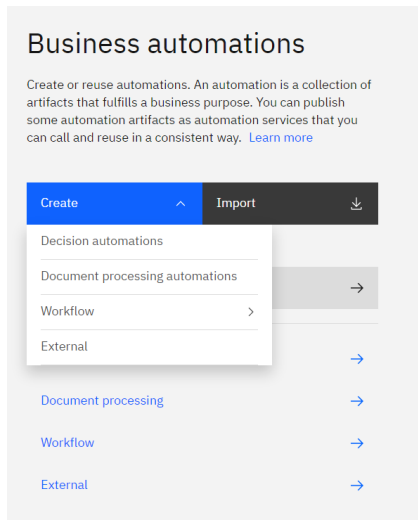
Create Import

Published automation services →

- Decision →
- Document processing →
- Workflow →
- External →

Published (3)	
WXC Skills Workflow	Published 09/05/2024
Client_Onboarding_Workflows Workflow	Published 09/05/2024
Client Onboarding Decisions Decision	Published 09/05/2024

\_4. Afin de créer un nouveau projet ADP, choisissez un projet de type « **Document Processing Automations** »



## Business automations

Create or reuse automations. An automation is a collection of artifacts that fulfills a business purpose. You can publish some automation artifacts as automation services that you can call and reuse in a consistent way. [Learn more](#)

Create Import

- Decision automations
- Document processing automations →
- Workflow >
- External

- Document processing →
- Workflow →
- External →

\_5. Assignez un nom à votre projet.

**Conseil important :** Vous partagez cet environnement avec d'autres personnes, utilisez votre prénom dans le nom de projet pour pouvoir le distinguer des projets des vos collègues.

Par exemple,

Create a document processing automation

Name

DMR-ISABELLE

Purpose (optional)

Describe the purpose of the document processing automation

Finalisez la création en appuyant sur le bouton

Create

Project initialization in progress

Your project is still initializing. Please wait until the initialization is complete.

\_6. Un peu de patience, et voilà votre premier projet est créé. Félicitations ! 😊

IBM Cloud Pak

[Business automations](#) /
 

## DMR-ISABELLE

Share

Last shared | a few se

Build

Enrich

Configure

<b>Document types and samples</b> Upload sample documents to define the types of documents you want the system to process.           →	<input checked="" type="checkbox"/> Ready	<b>3</b> types	<b>29</b> samples on average
<b>Classification model</b> Train the model to classify your documents.           →	<input checked="" type="checkbox"/> Ready	<b>3</b> types trained	<b>100%</b> accuracy
<b>Extraction model</b> Train the model to extract the data from your documents.           →	<input checked="" type="checkbox"/> Ready	<b>3</b> types trained	<b>97%</b> accuracy
<b>Data standardization</b> Map fields to new or existing data definitions.           →	<input type="checkbox"/> Not ready		
<b>Document retention</b> Determine how long you want documents to stay in your content repository.           →	<input checked="" type="checkbox"/> Ready	<b>3</b> types reviewed	



Vous pouvez en profiter pour parcourir les éléments de votre projet nouvellement créé.

Remarquez que celui-ci contient déjà quelques types de documents prédéfinis, un modèle de classification, et un modèle d'extraction.

Nous allons configurer les langues utilisées dans les modèles de classification, et d'extraction pour s'assurer de résultats optimaux.

\_7. Sélectionnez « **Configure** » dans la vue générale de votre projet

Business automations /

## DMR-ISABELLE

Build Enrich **Configure**

<b>Document types and samples</b> Upload sample documents to define the types of documents you want the system to process.	✓ Ready	2 types	28 samples on average
<b>Classification model</b> Train the model to classify your documents.	✓ Ready	2 types trained	100% accuracy

\_8. Sélectionnez le menu « **Langue Settings** »

**Language settings**

\_9. Activer le « Français », et désactivez les langues inutilisées.

Build Enrich **Configure** Last shared | 2 days ago

Import / Export project  
Document processing  
**Language settings**  
Git server configuration  
Webhook configuration

⚠ After you change the extraction language, you might need to change the language of natural language extractors, reanalyze training documents. Otherwise, it might impact accuracy.

### Language Settings

**Extraction language**  
Choose the language you want to use for extraction. You can choose more than one language, but this will impact accuracy. [Learn more](#)

Default  
☒ English

Other languages

☐ Dutch
 ☐ Portuguese (Brazil)
 ☒ French
 ☐ Spanish
 ☐ German




Pour maximiser la reconnaissance, il est important de sélectionner les langues susceptibles d'être rencontrées dans les documents, et de désélectionner les autres.

## Exercice 3 – Ajout d'une nouvelle classification

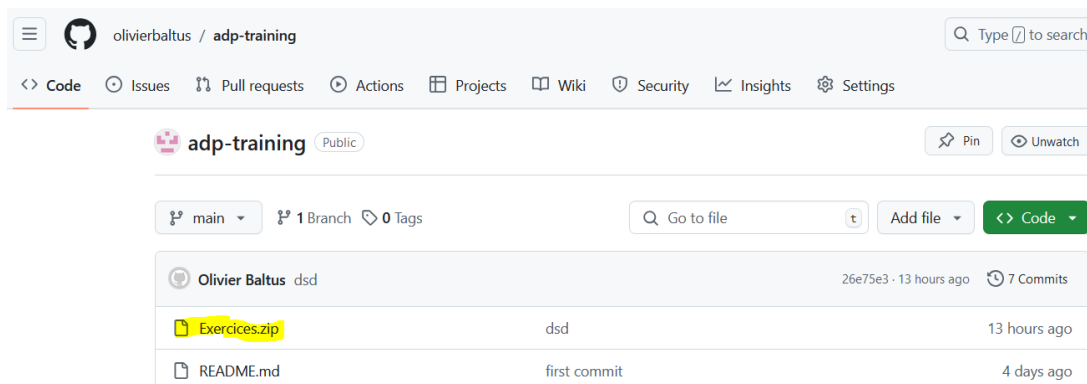
### Objectifs :

Dans cet exercice, vous allez apprendre à créer deux nouveaux types de document, entraîner ADP à les reconnaître.

### Prérequis:

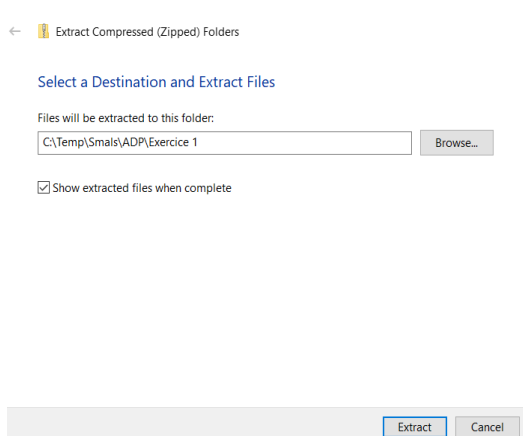
 Avant de commencer l'exercice, vous aurez besoin de documents pour configurer votre premier projet. Veuillez télécharger le fichier nécessaire depuis le GitHub à l'adresse :

<https://github.com/olivierbaltus/adp-training/tree/main>



\_1. Commencez par télécharger le fichier zip **Exercices.zip**, et décompressez son contenu dans le dossier de votre choix sur votre ordinateur.

Sélectionnez le fichier téléchargé, et cliquez sur le bouton de droite de la souris, et choisissez l'action « **Extract All** ».



\_2. Veuillez vous connecter à votre environnement, et ouvrir le projet que vous avez créé lors de l'exercice précédent.

IBM Cloud Pak

## Business automations

Create or reuse automations. An automation is a collection of artifacts that fulfills a business purpose. You can publish some automation artifacts as automation services that you can call and reuse in a consistent way. [Learn more](#)

Create Import

Published automation services →

Decision →

**Document processing** →

Workflow →

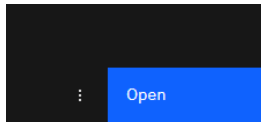
External →

Document processing automations (1)

**DMR-ISABELLE**

Last edited 09/06/2024

\_3. Cliquez sur le projet, et sélectionner « **Open/Ouvrir** »



Le contenu de votre projet devrait s'afficher.

IBM Cloud Pak

[Business automations](#) / **DMR-ISABELLE**

Build Enrich Configure

<b>Document types and samples</b> Upload sample documents to define the types of documents you want the system to process. →	<input checked="" type="radio"/> Ready	3 types	29 samples on average
<b>Classification model</b> Train the model to classify your documents. →	<input checked="" type="radio"/> Ready	3 types trained	100% accuracy
<b>Extraction model</b> Train the model to extract the data from your documents. →	<input checked="" type="radio"/> Ready	3 types trained	97% accuracy
<b>Data standardization</b> Map fields to new or existing data definitions. →	<input type="radio"/> Not ready		
<b>Document retention</b> Determine how long you want documents to stay in your content repository. →	<input checked="" type="radio"/> Ready	3 types reviewed	

\_4. Veuillez cliquer sur la carte «**Document types and samples** » pour accéder aux types de documents existants

## Document types and samples

Upload sample documents to define the types of documents you want the system to process.



Vous pouvez observer que votre projet contient déjà trois types de document à titre d'exemple :

## Document types and samples

Document types	
<b>Bill of Lading</b>	⋮
28 samples	
<b>Invoice</b>	
26 samples	
<b>Utility Bill</b>	
33 samples	

Puisque nous n'en avons pas besoin pour cet exercice, nous allons effacer ces trois types de document.

\_5. Veuillez effacer les trois types prédéfinis en cliquant sur le hamburger



Situé à côté de chacun des types. Et sélectionner l'action « **Delete** »

Document types		Document types		Document types	
<b>Bill of Lading</b>	⋮	<b>Invoice</b>	⋮	<b>Utility Bill</b>	⋮
28 samples		26 samples		33 samples	
Invoice		Utility Bill			
26 samples		33 samples			
Utility Bill					
33 samples					

Après avoir effacés les trois types, votre écran devrait ressembler à celui ci-dessous :



**You don't have any document types yet.**

Create new document types based on your organized sample documents. Include 5 samples for each type.

\_6. Nous allons maintenant définir notre premier type de document, que nous appellerons « **Facture** »

Cliquez sur le bouton « **Create Document type** » qui se trouve en haut, et à droite de l'écran

Create document type +

amples

Create document type +

#### Bill of Lading sample documents (28)

Review your existing and pre-trained document types. You can upload sample documents to refine each document type or create new document types. [Learn more](#)

Q Search sample documents

Upload ↗

<input type="checkbox"/>	Document name	↑	Status	↑	Date Added	↑	Added by	↑
<input type="checkbox"/>	<a href="#">BOL_005_1_1.1.pdf</a>		✓ Ready		Jun 28, 2022 11:06 PM		system	
<input type="checkbox"/>	<a href="#">BOL_005_2_1.pdf</a>		✓ Ready		May 26, 2023 8:48 AM		system	

Et sélectionnez « **With my sorted samples** », car nous allons fournir des échantillons pour notre nouveau type de document.



Create document type +

with my sorted samples

with my unsorted samples

without samples

## \_7. Remplissez les informations demandées :

Document type display name

7/50

Facture

Create document type

Document type display name

7/50

Facture

This is the name that will show up for you in the system. You can use characters from any language.

Document type symbolic name

7/50

Facture

This name will be used in code to identify the field type behind the scenes.

Classification confidence threshold %

70

-

+

Set a confidence level to be aware of documents that fall under the desired threshold. Documents under this threshold will show a warning.

Upload sample files (optional)

Maximum file size is 250 MB. Supported file types are .pdf, .doc, .docx and image formats.

Drag and drop files here or click to upload

☐ Fixed format ⓘ

Cancel


















Done

## \_8. Cliquez sur le lien en bleu

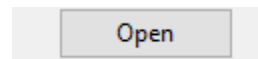
[Drag and drop files here or click to upload](#)

Pour déposer les échantillons correspondants au type « Facture » sur le serveur ADP.

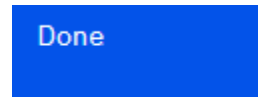
\_9. Sélectionnez tous les fichiers PDF contenus dans le dossier « **Training set/Facture** » résultant de la décompression de l'archive **Exercices.zip**

Exercices > Training set > Facture			
	0013497830 822400049149-1.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 242
	0013497830 822400102121-1.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 223
	0013497830 822400142205.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 240
	0013497830 822410025642-1.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 237
	0013497830 822410062499-1.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 316
	0013497830 822440036406-1.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 316
	0013497830 822460005768-5.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 192
	FACTURE-8.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 199
	FACTURE-10.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 200
	FACTURE-11.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 200
	FACTURE-12.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 200
	FACTURE-13.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 200
	FACTURE-14.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 201
	FACTURE-15.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 200
	FACTURE-16.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 200
	FACTURE-17.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 116
	FACTURE-18.pdf	07/09/2024 10:42	Adobe Acrobat Docum... 55

\_10. Appuyez sur le bouton « Open » pour valider le transfert des fichiers.



\_11. Terminez la création du type de document en appuyant sur le bouton « **Done** »

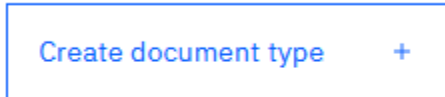


Veuillez noter que l'intégration des documents s'effectue en arrière plan par ADP.

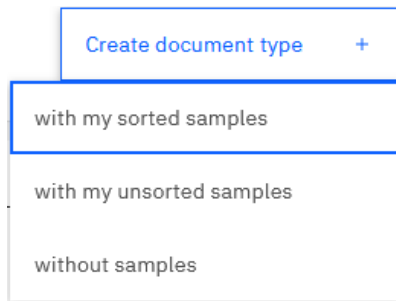
En attendant, vous pouvez consulter les échantillons pour vous familiariser avec leur contenu, ou bien éventuellement démarrer la création du second type de document.

\_12. Nous allons maintenant définir le deuxième type de document, et dans ce cas, il s'agira du type « **Justificatif de stationnement** »

Cliquez sur le bouton « Create Document type » qui se trouve en haut, et à droite de l'écran



Et sélectionnez « **With my sorted samples** », puisque nous allons également fournir des échantillons pour ce deuxième type de document.



\_13. Remplissez les informations demandées :

Create document type

Document type display name25/50

JustificatifStationnement

This is the name that will show up for you in the system. You can use characters from any language.

Document type symbolic name25/50

JustificatifStationnement

This name will be used in code to identify the field type behind the scenes.

Classification confidence threshold %

70

Set a confidence level to be aware of documents that fall under the desired threshold. Documents under this threshold will show a warning.

Upload sample files (optional)

Maximum file size is 250 MB. Supported file types are .pdf, .doc, .docx and image formats.

Drag and drop files here or click to upload

Fixed format

Feedback documents

Percentage of corrected documents to use in retraining

10













\_14. Cliquez sur le lien en bleu

[Drag and drop files here or click to upload](#)

Pour envoyer les échantillons correspondants au type « **JustificatifStationnement** » à ADP.

Exercices > Training set > Justificatif  
w)

\_15. Sélectionnez tous les fichiers PDF se trouvant dans le dossier « Training set/Justificatif »

 attach_18146818-2310130001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2311130001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2401020001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2401080001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2401300001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2402220001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2404220001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2405130001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2407180001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2407240001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2408070001.pdf	07/09/2024 10:42	Adobe Ac
 attach_18146818-2408070001-1.pdf	07/09/2024 10:42	Adobe Ac

\_16. Et appuyez sur le bouton « **Open** »

Open

\_17. Validez en pressant le bouton « **Done** »

Done

## Exercice 4 – Entraînement du modèle de classification

### Objectifs :

Dans cet exercice, nous allons entraîner le modèle de classification avec les nouveaux types documents.

\_1. Connectez vous à la plateforme IC4BA, et veuillez sélectionner votre projet en cliquant sur le lien :

[Business automations](#) / [DMR-ISABELLE](#) /

### Document types and samples

Document types

- Facture (26 samples)
- JustificatifStationnement (23 samples)

Facture sample documents (26)  
Review your existing and pre-trained document types. Y

Search sample documents

<input type="checkbox"/>	Document name
<input type="checkbox"/>	0013497830 822310078295.pdf
<input type="checkbox"/>	0013497830 822350040270.pdf

Vous pouvez observer que suite aux changements, le modèle de classification doit être entraîné à nouveau :

IBM Cloud Pak

Business automations / DMR-ISABELLE

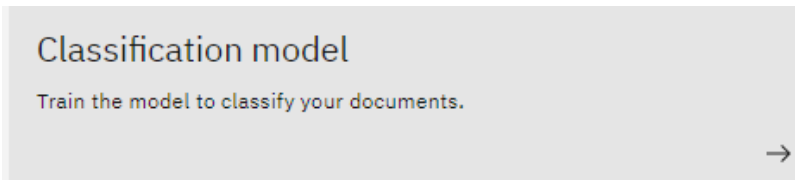
Build Enrich Configure

Component	Status	Count	Value
Document types and samples Upload sample documents to define the types of documents you want the system to process.	● Ready	2 types	25 samples on average
Classification model Train the model to classify your documents.	● Retrain	3 types trained	100% accuracy
Extraction model Train the model to extract the data from your documents.	● Retrain	3 types trained	97% accuracy
Data standardization Map fields to new or existing data definitions.	○ Not ready		
Document retention Determine how long you want documents to stay in your content repository.	● Ready	2 types reviewed	

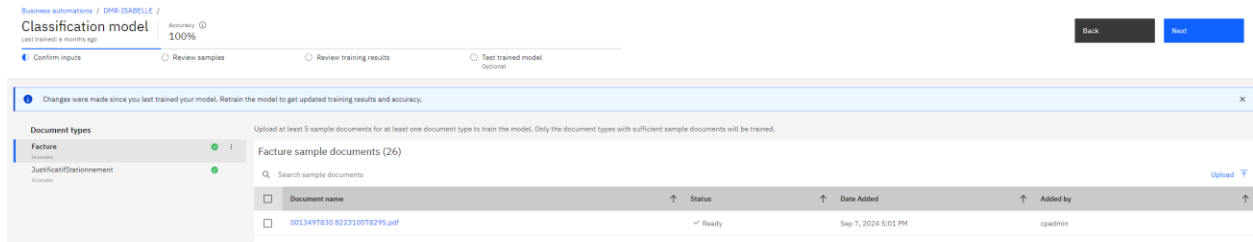


La plupart des changements de configuration ont impact au niveau des modèles, et nécessite un nouvel entraînement.

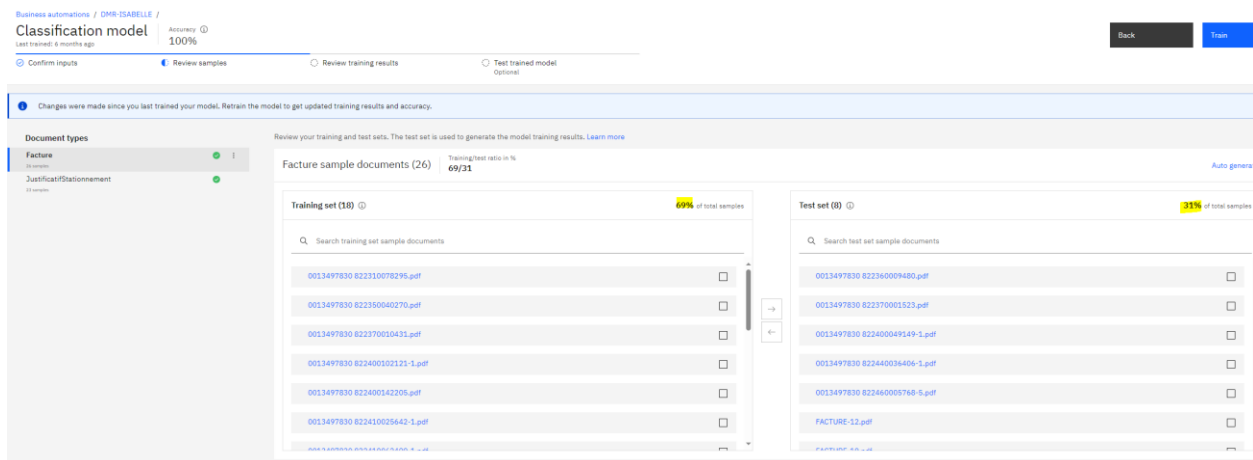
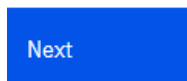
## \_2. Veuillez appuyer sur la carte "**Classification model** »



L'écran récapitulatif du modèle de classification s'affiche :



## \_3. Appuyez sur le bouton « **Next** »

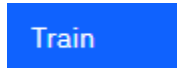


Veuillez observer qu'ADP répartit automatiquement les échantillons en deux jeux, un jeu d'entraînement, et un jeu de contrôle.

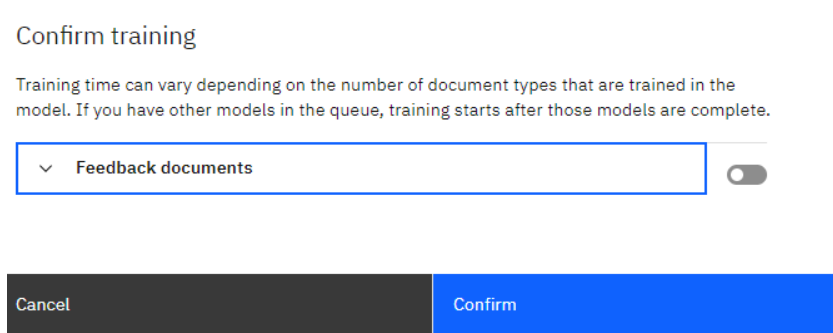
Vous avez la possibilité de modifier la clé de répartition ; le ratio 80/20 est généralement recommandé pour l'entraînement du modèle.

\_4. Nous allons maintenant démarrer l'entraînement du modèle de classification.

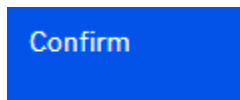
Veillez appuyer sur le bouton « **Train** » en haut à droite.



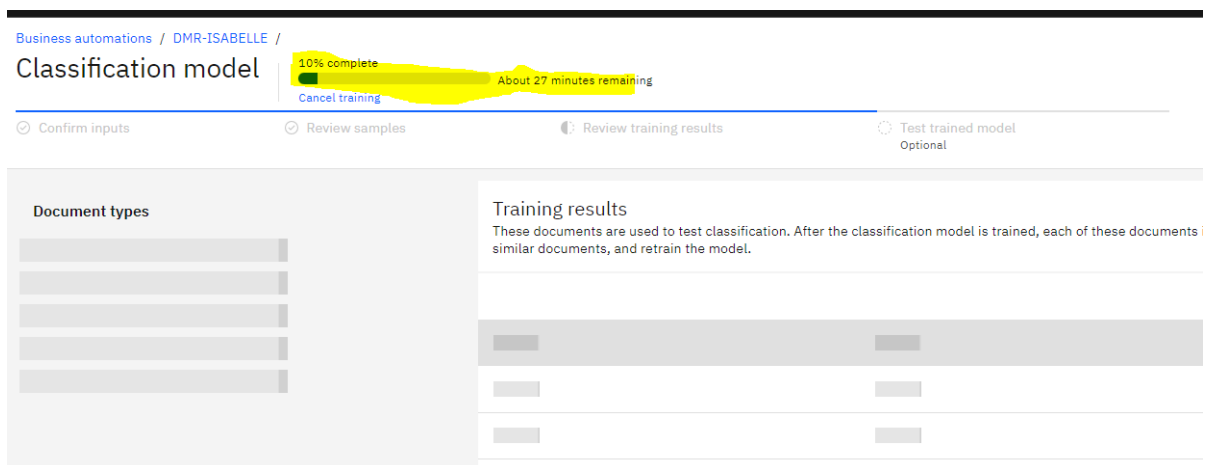
La boîte de dialogue suivante s'affiche :



\_5. Confirmez l'entraînement en appuyant « **Confirm** »



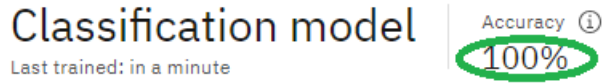
L'écran de progression suivant s'affiche,



Veillez noter que l'entraînement du modèle de classification se déroule en arrière-plan, et peut être interrompu à tout moment.



Lorsque l'entraînement du modèle de classification est terminé, ADP affiche un niveau de précision global



Et, pour chacun des types de classification, un tableau récapitulatif avec les résultats de la classification.

C'est-à-dire, le résultat de classification, et le niveau de confiance.

Document types	Training results		
Facture	These documents are used to test classification. After the classification model is trained, each of these documents is tested to see whether the system can correctly determine the document type. If a document has an incorrect classification result or a confidence warning, review the inputs, and similar documents, and retrain the model.		
Justificatif/Stationnement	Q Search sample documents		
Document	Classified as	Classification result	Confidence
0013497830 822360009480.pdf	Facture	Correct	94.46%
0013497830 822370001523.pdf	Facture	Correct	94.79%
0013497830 822400049149-1.pdf	Facture	Correct	95.02%
0013497830 823480036486-1.pdf	Facture	Correct	95.01%
0013497830 823460005765-6.pdf	Facture	Correct	94.85%
FACTURE-12.pdf	Facture	Correct	95.02%
FACTURE-19.pdf	Facture	Correct	94.99%
FACTURE-22.pdf	Facture	Correct	95.01%



Un niveau de confiance supérieur à 80% est généralement considéré comme acceptable.



## Exercice 5 – Validation du modèle de classification

### Objectifs :

Dans cet exercice, nous allons tester notre nouveau modèle de classification.

\_1. Connectez-vous à la plateforme IC4BA, et veuillez ouvrir votre projet ADP.

[Business automations](#) /

DMR-ISABELLE

Build	Enrich	Configure
<b>Document types and samples</b> Upload sample documents to define the types of documents you want the system to process.	<input checked="" type="radio"/> Ready	2 types 28 samples on average
<b>Classification model</b> Train the model to classify your documents.	<input checked="" type="radio"/> Ready	2 types trained 100% accuracy
<b>Extraction model</b> Train the model to extract the data from your documents.	<input checked="" type="radio"/> Retrain	3 types trained 97% accuracy
<b>Data standardization</b> Map fields to new or existing data definitions.	<input type="radio"/> Not ready	
<b>Document retention</b> Determine how long you want documents to stay in your content repository.	<input checked="" type="radio"/> Ready	2 types reviewed

\_2. Sélectionnez le modèle de classification en cliquant sur

**Classification model**  
 Train the model to classify your documents.

→

\_3. A partir de l'écran de classification,

IBM Cloud Pak

Business automations / DMR-ISABELLE / Classification model

Accuracy ⓘ 100%

Last trained: an hour ago

Confirm inputs Review samples Review training results Test trained model (Optional)

Document types

Document type	Samples	Status
Facture	32 samples	✓
JustificatifStationnement	23 samples	✓

Upload at least 5 sample documents for at least one document type to train the model. Only the document types with sufficient samples will be used for training.

Facture sample documents (32)

Search sample documents

veuillez sélectionner « **Test trained model** »

Test trained model  
Optional

\_4. A partir du dossier contenant les échantillons, sélectionnez différents fichiers de test contenus dans le dossier « Testing Sets »

Exercices > Testing set

Optional

Document types

Document type
Facture
JustificatifStationnement

Testing results

Review how accurately the model tests your sample documents.

You haven't uploaded any documents yet.

Upload 5-10 new documents that represent your different document types. This can be a mix of different formats (.pdf, .doc, .docx) and image formats. Click Upload to start uploading documents.

Upload

Et observez, le résultat de la classification, et le taux de confiance associé en fonction du type de document, et de l'échantillon de test choisi.

Pour le type « facture »,

Document types		Testing results				
Facture		Review how accurately the model tests your sample documents.				
JustificatifStationnement		<div> <div>Q</div> <div>Search sample documents</div> <div>Start over ↺ Upload ↗</div> </div>				
Document	↑	Classified as	↑	Classification result	↑	Confidence
✓ 1_unknown.pdf		Facture		Correct		63.09%
✓ 1_unknown.pdf		Facture		Correct		63.09%
✓ 2_unknown.pdf		Facture		Correct		69.67%
✓ 3_unknown.pdf		Facture		Correct		93.47%

Pour le type « Justification Stationnement »,

Document types		Testing results				
Facture		Review how accurately the model tests your sample documents.				
JustificatifStationnem...		<div> <div>Q</div> <div>Search sample documents</div> <div>Start over ↺ Upload ↗</div> </div>				
Document	↑	Classified as	↑	Classification result	↑	Confidence
✓ 1_unknown.pdf		Facture		Incorrect		63.09%
✓ 2_unknown.pdf		Facture		Incorrect		69.67%
✓ 3_unknown.pdf		Facture		Incorrect		93.47%



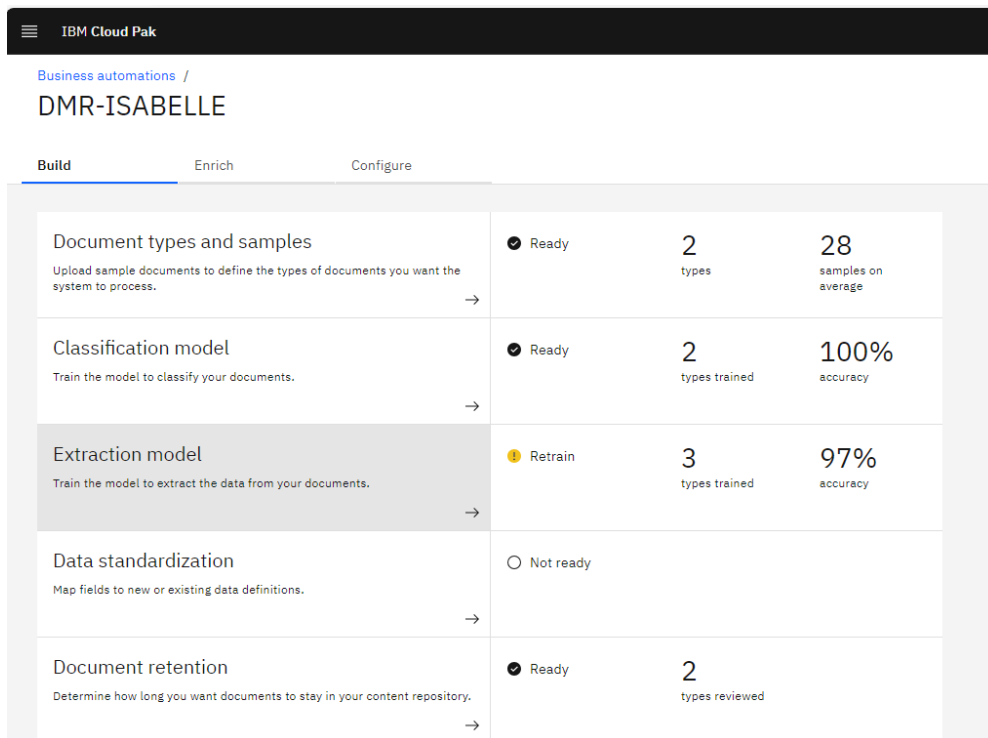
On observe différents cas de figure, des documents qui sont classés de manière incorrecte avec un taux de confiance très bas (<70%). Dans ce cas, on considère que le résultat de la classification n'est pas correct. Un document qui est reconnu correctement avec un taux de confiance très élevé (>90%).

## Exercice 6 – Modèle d'extraction

### Objectifs :

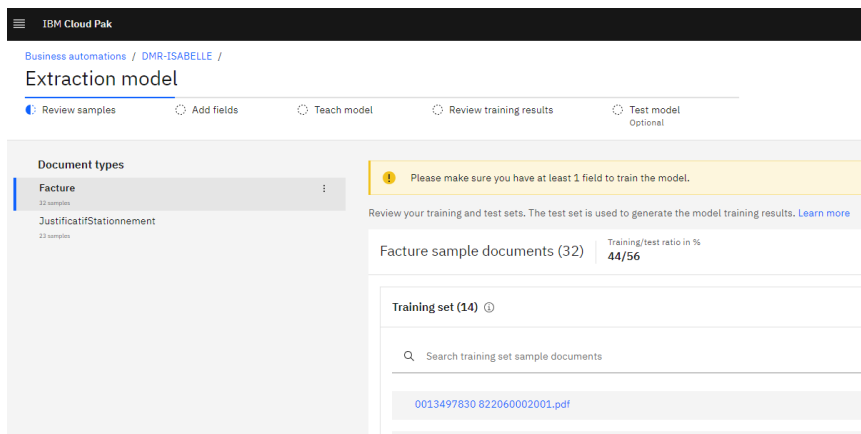
Dans cet exercice, vous allez apprendre à définir de nouveaux champs, et à réaliser la configuration de l'extraction dans ADP.

\_1. Connectez vous à la plateforme IC4BA, et veuillez ouvrir votre projet ADP.



Step	Status	Types	Accuracy
Document types and samples Upload sample documents to define the types of documents you want the system to process.	Ready	2 types	28 samples on average
Classification model Train the model to classify your documents.	Ready	2 types trained	100% accuracy
Extraction model Train the model to extract the data from your documents.	Retrain	3 types trained	97% accuracy
Data standardization Map fields to new or existing data definitions.	Not ready		
Document retention Determine how long you want documents to stay in your content repository.	Ready	2 types reviewed	

\_2. Sélectionnez la carte modèle d'extraction « **Extraction Model** »



**Extraction model**

Review samples | Add fields | Teach model | Review training results | Test model (Optional)

**Document types**

- Facture (32 samples)
- JustificatifStationnement (23 samples)

Please make sure you have at least 1 field to train the model.

Review your training and test sets. The test set is used to generate the model training results. [Learn more](#)

Facture sample documents (32) | Training/test ratio in % 44/56

Training set (14) @

Search training set sample documents

0013497830 822060002001.pdf

\_3. Nous allons maintenant ajouter deux nouveaux champs pour le type de document « Facture »

- Numéro de client (chaîne de caractères, obligatoire)
- Numéro de facture (chaîne de caractères, obligatoire)

Cliquez sur le lien en bleu « **Add fields** » pour afficher la page d'ajout de nouveau champ.

Business automations / DMR-ISABELLE /

## Extraction model

[Review samples](#) [Add fields](#) [Add fields](#)

---

Business automations / DMR-ISABELLE /

### Extraction model

[Review samples](#) [Add fields](#) [Teach model](#) [Review training results](#) [Test model Optional](#)

**Document types**

- Facture**
- JustificatifStationnem...

**Fields to extract for Facture (0)**  
Add attributes that you want to extract from this document type. [View a sample document](#)

**No fields added**  
Add fields that you want to extract from this document  
You will look for these fields when you annotate sample

[Add fields](#) +

[Create composite](#)

[Create table](#)

\_4. Cliquez à nouveau sur « Add fields » en bleu



**No fields added**

Add fields that you want to extract from this document type.

You will look for these fields when you annotate sample documents

[Add fields](#) +

[Create composite](#)

[Create table](#)

\_5. Nous allons d'abord ajouter le champ « Numéro de Client »,

Field display name 16/50

This is the name that will show up for you in the system. You can use characters from any language.

et nous allons indiquer que ce champ est obligatoire.

☒ **This field is required** ⓘ

\_6. Ensuite, nous validons la création du champ en cliquant sur le bouton bleu « **Add field** »

[Add field](#)

Nous venons de créer un nouveau champ à notre type de document Facture.

Fields to extract for Facture (1)

Add attributes that you want to extract from this document type. [View a sample document](#)

Q

Search fields

Add table

Add composite

Add fields

+

Field name	Value format	Field type	Required	Sensitive	Date created	
Numéro de Client	abc Text	String	Yes	No	Sep 8, 2024 3:07 PM	<div><div></div><div></div></div>



\_7. Nous allons maintenant ajouter le champ « Numéro de Facture »,

Add a new field

Add a new field to extract from the document type Facture

\_8. Validons la création du champ en cliquant sur le bouton bleu « **Add field** »

Add field

Nous venons de créer un deuxième champ à notre type de document Facture.

\_9. Nous pouvons maintenant apprendre à ADP à reconnaître les numéros de client et de facture sur les documents.

Appuyez sur le bouton bleu en haut à droite

Next

\_10. Sélectionnez ensuite le bouton « **Teach samples** »

Teach samples →

\_11. Nous allons maintenant indiquer où se trouve notre champ numéro de client sur le document.

Nous commençons par identifier le libellé du champ. « **Field Label** »

Cliquez sur le bouton « **Draw** » au niveau de « **Field Label** »

Field label (optional)

Drawing...



Captured field label



\_12. Identifiez le libellé du champ « Numéro du client » à extraire au niveau du document

TVA: 0812806748

# Facture

## Vos références

Numéro du client	Votre code VOO	Nu
0013497830	28F-EC4-7BA	82:

\_13. Validez la sélection en cliquant sur « **Save Selection** »

Save selection

\_14. Exécutez la même opération pour identifier la valeur du numéro de client sur la facture.

Cliquez sur « **Drawing** » au niveau de « Field Value »

Field value

Drawing...

Captured field value

\_15. Identifiez la zone en l'encadrant à l'aide du lasso.

# Facture

## Vos références

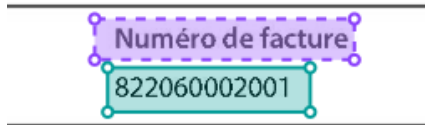
Numéro du client
0013497830

\_16. Sauvez la sélection du champ, en cliquant sur le bouton bleu « **Save Selection** »,

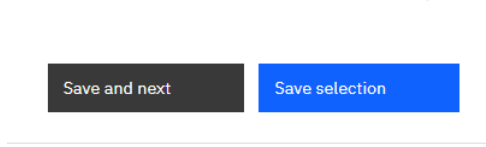
Save and next

Save selection

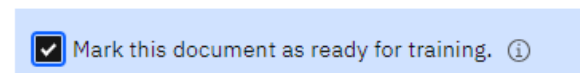
\_17. Répétez les mêmes opérations pour identifier le numéro de facture sur le client



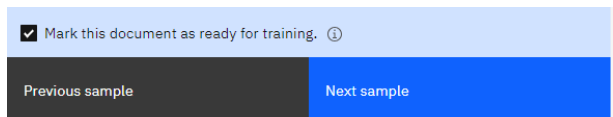
\_18. Sauvez la sélection du champ, en cliquant sur le bouton bleu « **Save Selection** »,



\_19. N'oubliez d'indiquer à ADP que ce document est prêt pour l'entraînement en cliquant  
« **Mark this document as ready for training** »



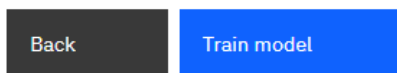
\_20. Vous pouvez passez à l'apprentissage du document suivant en cliquant sur le bouton bleu  
« **Next Sample** »



\_21. Répétez l'apprentissage pour tous les échantillons. Quand il n'y a plus d'échantillons, vous pouvez retourner dans l'écran de configuration de l'extraction, en appuyant sur « **Back** » en haut à gauche.



\_22. Vous pouvez maintenant entrainer le modèle d'extraction, en cliquant sur le bouton « **Train Model** »



\_23. L'écran de configuration de l'entraînement s'affiche, laissez les options par défaut, et appuyer sur le bouton bleu « **Confirm** »

Confirm training

Training the model can take several hours or multiple days depending on how many document types you are training. If you have other models in the queue, training starts after those models are complete.

☐ Fast training ☐

☐ Feedback documents ☐

Et patientez...

\_24. A la fin de l'entraînement, le résultat de l'extraction s'affiche.

IBM Cloud Pak

Business automations / DMR-ISABELLE /

Extraction model

☐ Review samples
 ☐ Add fields
 ☐ Teach model
 ☒ Review training results
 ☐ Test model Optional

Last training results | Previous training results | Last saved model

- average accuracy	- average table accuracy	100% average confidence	2 fields	11 samples tested
--------------------	--------------------------	-------------------------	----------	-------------------

Verify results

Review below document type(s) to verify training results and improve your model accuracy.

Document type	Fields	Samples tested	Missing required fields	Average fields found	Average confidence
Facture	2	11	0	2	100%

\_25. Vous pouvez maintenant tester le résultat de l'extraction en cliquant sur le « **Document type** » Facture

Facture

\_27. Le résultat de l'extraction s'affiche pour le type de document « **Facture** »

Business automations / DMK-LSABELLE /

Extraction model

Back

Next

Review samples

Add fields

Teach model

Review training results

Test model  
Optional

Last training results

Previous training results

Last saved model

← Back to all results

Facture

Verify all

For the document type(s) selected in Verify results step, review below testing samples to establish the ground truth values.

Testing sample	↑	Fields found	Average confidence	Status	Action
0013497830 822140018738-1.pdf		2	100%	Needs verification	<a href="#">Start</a>
0013497830 822150059412.pdf		2	100%	Needs verification	<a href="#">Start</a>
0013497830 822310078295.pdf		2	100%	Needs verification	<a href="#">Start</a>
0013497830 822350040270.pdf		2	100%	Needs verification	<a href="#">Start</a>
0013497830 822360009480.pdf		2	100%	Needs verification	<a href="#">Start</a>