

Auteurs : BARAL Mathis – LEPRAT Florian - TUVATALENI Bem Amado

CONTEXTE ET OBJECTIF

Fondée en 2000, @GP est une société spécialisée dans le développement de solutions de dématérialisation de documents pour les entreprises. Face à la généralisation de la facture électronique, @GP a identifié le besoin d'automatiser le processus d'extraction et de qualification des données à partir de factures PDF.

Actuellement, cette tâche est réalisée manuellement, nécessitant en moyenne 2 heures par opérateur pour créer des modèles graphiques d'extraction.

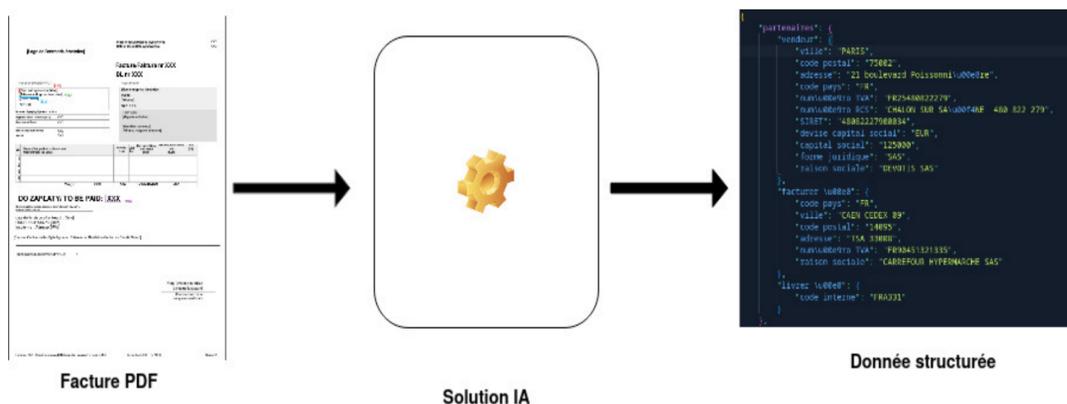
L'objectif de ce projet est donc de concevoir et développer un système basé sur l'intelligence artificielle (IA) capable de :

- Traiter et lire des factures PDF textuelles.
- Extraire automatiquement les informations clés des factures, telles que le nom de l'entreprise, l'adresse, le montant total, etc.
- Qualifier ces informations pour les rendre exploitables par les systèmes de gestion des entreprises clientes.
- Gérer un grand nombre de formats de factures différents avec un haut niveau de précision et de performance.



MÉTHODES ET DÉVELOPPEMENTS

Après avoir exploré différentes approches d'IA (classifieurs multi-classes, BERT, LLMs, LayoutLM), l'équipe a opté pour le modèle LiLT. Ce choix s'explique par les performances accrues de ce modèle dans l'extraction d'informations à partir de documents visuels tels que les PDF, ainsi que par ses conditions de licence avantageuses.

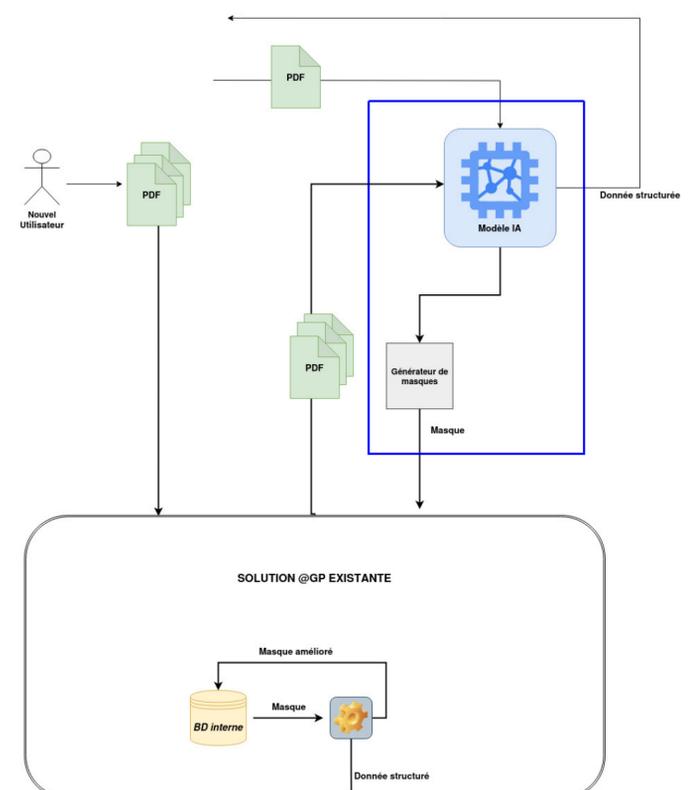


Étapes clés du développement :

- Collecte et préparation des données
- Entraînement du modèle LiLT
- Développement de l'API REST
- Intégration et déploiement

RÉSULTATS ET CONCLUSION

- Le modèle LiLT a démontré une capacité d'extraire une bonne partie des informations présentes dans les factures PDF.
- La solution basée sur l'IA pourra être utilisée pour assister le processus d'extraction d'informations et de génération de masques d'extraction à partir de factures PDF. Pouvant réduire le temps nécessaire pour réaliser cette tâche manuellement.
- L'architecture basée sur Docker facilitera l'intégration et le déploiement du système dans l'infrastructure du client.



MOTS-CLÉS : Machine learning, EDI, IA, Docker