Synthèse du Rapport de Stage Développement d’un Portail RH avec Intelligence

Artificielle

Hamza Elmourabit 31 juillet 2025

# Contexte du Stage

Ce stage d’initiation s’est déroulé au sein de l’Office National de l’Électricité et de l’Eau Potable (ONEE) – Branche Électricité, dans le cadre du stage pour la filière Big Data et Systèmes d’Information de l’ENSA Berrechid. D’une durée d’ un mois, il avait pour objectif principal de développer un module d’intelligence artificielle pour l’extraction automatique de données à partir de documents administratifs (CIN, diplômes, CV) afin d’optimiser les processus RH.

# Problématique

Le département RH de l’ONEE faisait face à des défis majeurs dans la gestion documentaire :

* Processus manuel chronophage (15-20 minutes par document)
* Risque élevé d’erreurs humaines lors de la saisie
* Difficultés d’accès et de recherche des informations
* Manque de standardisation des données
* Problématiques de sécurité et de confidentialité

Ces limitations créaient un goulot d’étranglement significatif dans les processus RH et limitaient l’efficacité opérationnelle du département.

# Objectifs du Projet

* **Objectif principal** : Développer un module IA capable d’extraire automatiquement les données clés des documents administratifs

### Objectifs spécifiques :

* + Maîtrise des technologies OCR (Tesseract)
  + Extraction automatisée des champs pertinents
  + Structuration des données en format JSON
  + Intégration dans des bases de données (MongoDB/MySQL)
  + Développement d’une interface utilisateur intuitive

# Solution Développée

## Architecture Technique

Le projet a été implémenté selon une architecture modulaire et scalable :

Document

Prétraitement (OpenCV)

OCR

(Tesseract)

Extraction (Regex/NLP)

Interface (Streamlit)

JSON

Stockage (MongoDB/MySQL)

Figure 1 – Architecture du système d’extraction de données

## Technologies Utilisées

|  |  |
| --- | --- |
| **Composant** | **Technologie** |
| Langage principal | Python 3.10+ |
| Traitement d’images | OpenCV |
| OCR | Tesseract (support arabe/français) |
| Extraction de données | Regex, dictionnaires, NLP |
| Stockage | MongoDB (NoSQL) + MySQL (relationnel) |
| Interface utilisateur | Streamlit (prototype) |
| Backend | Flask |
| Déploiement | Docker |

Table 1 – Stack technologique utilisé

# Résultats Obtenus

## Performances du Système

Le module d’extraction a été évalué sur un échantillon de 100 documents :

|  |  |  |
| --- | --- | --- |
| **Type de document** | **Taux de réussite** | **Temps moyen** |
| CIN | 92% | 3.2 secondes |
| Diplômes | 87% | 4.5 secondes |
| CV | 75% | 6.8 secondes |

Table 2 – Performances par type de document

## Gains Opérationnels

L’implémentation du système a généré des améliorations significatives :

100

85

70

75

60

80

Amélioration (%)

60

40

20

0

Temps Erreurs AccessibilSi **t**éandardisation

Métrique

Figure 2 – Gains opérationnels obtenus

* **Réduction de 70%** du temps de traitement des documents
* **Diminution de 85%** des erreurs de saisie
* **Amélioration de 60%** de l’accessibilité des données
* **Standardisation de 75%** des processus RH

# Difficultés Rencontrées et Solutions

|  |  |  |
| --- | --- | --- |
| **Difficulté** | **Description** | **Solution apportée** |
| Qualité des documents | Documents scannés de faible  qualité, ombres, flou | Algorithmes avancés de pré-  traitement avec OpenCV |
| Reconnaissance arabe | Complexité de l’écriture  arabe, polices variées | Configuration optimisée de  Tesseract, post-traitement |
| Variabilité des formats | Structures hétérogènes des  CV | Approche hybride combinant  règles et NLP |
| Gestion des erreurs | Erreurs d’OCR, champs man- quants | Système de scoring de  confiance, interface de correc- tion |

Table 3 – Principales difficultés et solutions

# Compétences Développées

Ce stage a permis de développer et de renforcer de nombreuses compétences :

## Compétences Techniques

* Maîtrise approfondie de Python et de son écosystème data science
* Expertise en traitement d’images (OpenCV) et OCR (Tesseract)
* Conception de bases de données hybrides (MongoDB + MySQL)
* Développement d’interfaces web avec Streamlit et Flask
* Implémentation d’algorithmes d’extraction et de NLP

## Compétences Professionnelles

* Gestion de projet et méthodologie agile
* Analyse des besoins et conception de solutions techniques
* Résolution de problèmes complexes
* Communication technique et présentation
* Adaptation à l’environnement professionnel

# Apports à l’ONEE

Le projet a apporté des bénéfices tangibles à l’organisation :

* **Optimisation des processus RH** : Automatisation des tâches répétitives, gain de temps significatif
* **Amélioration de la qualité des données** : Réduction des erreurs, standardisation des infor- mations
* **Modernisation digitale** : Positionnement comme acteur innovant dans la transformation numérique
* **Gain de productivité** : Réaffectation du personnel RH vers des tâches à plus forte valeur ajoutée

# Perspectives d’Amélioration

Plusieurs axes d’amélioration ont été identifiés pour l’évolution du projet :

### Intégration de modèles IA avancés :

* + Utilisation de LayoutLM pour la compréhension de documents
  + Fine-tuning sur des données marocaines spécifiques
  + Intégration de modèles de langage (BERT, GPT)

### Industrialisation de la solution :

* + Architecture microservices pour meilleure scalabilité
  + Création d’API REST pour l’intégration système
  + Mise en place d’un système de monitoring avancé

### Extension fonctionnelle :

* + Support de nouveaux types de documents
  + Amélioration continue des taux d’extraction
  + Développement de fonctionnalités analytiques avancées

# Conclusion

Ce stage a été une expérience enrichissante qui a permis de mettre en pratique les connaissances académiques dans un contexte professionnel concret. Le développement du portail RH avec intégration d’IA a démontré le potentiel considérable de l’intelligence artificielle pour transformer la gestion des ressources humaines.

Les résultats obtenus dépassent les attentes initiales, avec des gains opérationnels significatifs et une amélioration notable de l’efficacité des processus RH. Ce projet constitue une base solide pour la transformation numérique continue de l’ONEE et peut servir de modèle pour d’autres organisations.

Les compétences techniques et professionnelles acquises durant ce stage représentent un atout majeur pour mon futur parcours professionnel dans le domaine de la data science et de l’intelligence artificielle