

Semaine 2 : Développement du moteur IA (NLP et Classification)

(du 16 au 20 juin 2025)

Semaine 2 : Développement du moteur d'Intelligence Artificielle

❖ Objectif principal :

Développer le moteur d'IA composé :

- D'un module NLP pour extraire les informations des projets ;
- D'un modèle de classification pour attribuer chaque projet à l'un de différents domaines

❖ Tâches à réaliser

1. Collecte des données de projets

- Rassembler les exemples de projets soumis
- Nettoyer les données en enlevant les métadonnées
- Extraire le texte brut avec python

🚦 Concernant les données de projets soumis, nous allons créer un dataset synthétique réaliste de 30 à 50 projets ayant tous la même structure qui suit :

- Titre du projet
- Date de soumission
- Porteur du projet
- Destinataire
- Résumé du projet
- Objectifs du projet
- Description du projet
- Localisation
- Bénéficiaires
- Budget prévisionnel
- Conclusion

2. Nettoyage et Vectorisation

- Normalisation du texte
- Vectorisation avec TF-IDF

3. Extraction NLP

- Identification automatique avec Spacy ou Bert

4. Développement du modèle de classification

- Choix du modèle : RandomForest ou SVM
- Etiquetage manuel (si données insuffisantes)
- Séparation en dataset d'entraînement / validation
- Entraînement du modèle
- Validation croisée de la précision (f1_score, accuracy)

5. Evaluation et Tests

Créer un script de prédiction : Lorsqu'on lui donne un document, il renvoie :

- Le domaine
- Le résumé des informations extraites après analyse
- Le score de confiance

❖ Outils à utiliser :

- Vs Code IDE
- Python et certaines librairies
- Streamlit (pour certains tests qui seront réalisés)

Livrables attendus en fin de semaine

- Moteur NLP fonctionnel
- Modèle de classification entraîné
- Scripts d'entraînement et de prediction

Répartition par membre

Membre de l'équipe	Tâches principales assignées
Léa Botoko, Jonas Tshianyi, Soleil Kadima	<ol style="list-style-type: none">1. Collecte des données de projets2. Préparation de données
Daniel Tshibangu, Jenovic Kabongo	<ol style="list-style-type: none">1. Extraction NLP2. Développement du modèle de classification3. Evaluation et Tests