

**LIVRET**

**D’APPRENTISSAGE**

[NOM Prénom]

**Promotion [...]**

**Certification RNCP38616**

➨ Bloc 03 : Appliquer des techniques d’analyse IA via des algorithmes d'apprentissage automatiques.

➨ Bloc 05 : Développer et mettre en production des algorithmes d’IA par apprentissage profond.

**Avant de commencer ….**

**Bienvenue,**

**Voici le *template* pour l’élaboration de votre Livret d’apprentissage.**

**Ne changez pas sa mise en forme, afin que les examinateurs concentrent leur attention sur vos réponses.**

**Respectez la taille du texte, et l’espace disponible.**

**Savoir être précis et concis est la clé !**

***Daniel Villa Monteiro***

**CAS PRATIQUE**

**Epreuve 1**

📅 Semaine 5

**Modalités d’évaluation :**

Sur la base d’un cas d’entreprise réel ou fictif, le/la candidat(e) doit identifier les enjeux/problématiques rencontrées par l’entreprise. Il/elle doit traduire les enjeux du client en objectifs réalisables. A partir de ces objectifs, il/elle doit programmer, entraîner et utiliser un modèle d’apprentissage profond.

**💡 Cette première épreuve du Livret d’apprentissage permet d’évaluer les compétences suivantes :**

➨Bloc 3 : Appliquer des techniques d’analyse IA via des algorithmes d’apprentissage automatique

🎓 C1 : **Sélectionner l’algorithme d’apprentissage le plus adapté** en comparant les performances et les caractéristiques des différentes familles d’algorithmes afin d'apporter une réponse pertinente à la problématique métier rencontrée.

🎓 C2 : **Préparer et transformer des données** en utilisant des techniques de prétraitement (preprocessing) pour les adapter aux spécificités du modèle d'apprentissage automatique choisi.

🎓 C3 : **Entraîner un modèle d'apprentissage automatique** en optimisant une loss function (fonction de coût) à partir des données d’entraînement afin de permettre à l’algorithme d’effectuer le moins d’erreurs possibles selon des indicateurs de succès clairement définis.

**Epreuve 1 - 1ère Partie**

**Analyse du besoin :**

|  |
| --- |
| L’objectif de ce projet est d’automatiser la catégorisation d’offres de services issues d’une plateforme freelance (type Fiverr) selon plusieurs critères métiers : la thématique de l’offre, le niveau de compétence requis et une estimation de la fourchette de prix. Le besoin métier exprimé consiste à aider une entreprise à évaluer rapidement la position d’une nouvelle offre dans un référentiel marché, afin de gagner du temps dans la veille concurrentielle.chéma |

**Etat de l’art :**

|  |
| --- |
| Des interfaces utilisateurs d’IA sont souvent implémentées avec des frameworks comme Streamlit, Gradio ou des solutions plus complètes (Flask/Django + frontend). Gradio s’impose ici comme un choix rapide et accessible pour intégrer un modèle ML dans une interface de démonstration. |

**Choix technique :**

|  |
| --- |
| Nous avons retenu le pipeline suivant : - TF-IDF sur le titre pour extraire les features textuelles - OneHotEncoder sur les variables catégorielles (niveau de service notamment) - XGBoostClassifier pour la classification (performant et robuste aux données déséquilibrées) Les modèles sont validés par cross-validation, optimisés avec GridSearchCV, puis sauvegardés (pickle).héma |

**Epreuve 1 - 2ème Partie**

**Mise en oeuvre technique :**

|  |
| --- |
| Nettoyage des données (NaN, encodage, harmonisation des catégories) - Création d’un ColumnTransformer avec pipeline textuel et pipeline catégoriel - Entraînement de plusieurs modèles de classification (XGBoost, Random Forest, etc.) - Sélection du meilleur modèle sur la base des scores de accuracy et de f1-score - Sauvegarde du modèle entraîné et du préprocesseur (joblib) pour réutilisation - Tests sur nouveaux échantillons pour valider la cohérence éma |

**Epreuve 1 - 3ème Partie**

**Bilan :**

|  |
| --- |
| Le modèle final (XGBoost) atteint un score d’accuracy supérieur à 87 % sur les classes principales. La vectorisation des titres via TF-IDF et la combinaison avec des données tabulaires offre une bonne complémentarité. |

**Amélioration :**

|  |
| --- |
| - Améliorer la gestion des classes déséquilibrées (ex. rareté des offres haut de gamme) - Ajouter de l’analyse sémantique via des embeddings type Word2Vec ou BERT - Mettre en place une pipeline de validation automatique à chaque ajout de données |

**PROJET FINAL**

**Epreuve 2**

📅 Semaine 13

**Modalités d’évaluation :**

A partir d’un cas d’entreprise réelle ou fictive, le/la candidat(e) doit développer une application exploitable par un client final intégrant des solutions IA.

**💡 Cette seconde épreuve du Livret d’apprentissage permet d’évaluer les compétences suivantes :**

➨Bloc 5 : Développer et mettre en production des algorithmes d’IA par apprentissage profond (Deep Learning)

🎓 C1 : **Préparer des données non structurées** en les convertissant en données numériques et sous forme tabulaires pour servir de données d'entraînement à un algorithme d’apprentissage profond.

🎓 C2 : **Sélectionner l’algorithme d’apprentissage profond le plus adapté** en comparant les performances et les caractéristiques des différentes familles d’algorithmes afin d'apporter une réponse pertinente adaptée à la problématique métier rencontrée.

🎓 C3 : **Entraîner un modèle d'apprentissage profond** en optimisant une loss function (fonction de coût) à partir des données d’entraînement afin de permettre à l’algorithme d’effectuer le moins d’erreurs possibles selon des indicateurs de succès clairement définis.

🎓 C4 : **Déployer efficacement un modèle d'apprentissage profond** en utilisant des outils et plateformes de production adaptés (MLOps), pour assurer une accessibilité et une performance optimale des prédictions de l'algorithme aux utilisateurs finaux.

**Epreuve 2 - 1ère Partie**

***Compréhension du besoin client***

**Décrivez les besoins de votre client fictif, auquel votre projet tente de répondre :**

|  |
| --- |
| Le client souhaite disposer d’une application Web simple permettant de prédire automatiquement les caractéristiques d’une offre de service freelance (catégorie, niveau, tranche de prix), à partir d’un formulaire. L’objectif est de permettre à tout collaborateur de saisir une offre, même sans connaissance technique. |

**Epreuve 2 - 1ère Partie**

***Etat de l’art***

**Décrivez l’état de la concurrence et des recherches scientifiques quant à votre sujet d’application :**

|  |
| --- |
| Des interfaces utilisateurs d’IA sont souvent implémentées avec des frameworks comme Streamlit, Gradio ou des solutions plus complètes (Flask/Django + frontend). Gradio s’impose ici comme un choix rapide et accessible pour intégrer un modèle ML dans une interface de démonstration. |

**Epreuve 2 - 1ère Partie**

***Traduction et choix technique du projet***

**Expliquer l’ensemble de la stack technique que vous utilisez dans votre projet :**

|  |
| --- |
| - Backend IA : pipeline preprocessor.pkl + xgb\_model.pkl - Frontend : application Gradio, formulaire texte + dropdown - Environnement Python : scikit-learn, xgboost, pandas, gradio - Hébergement : exécution en local, déploiement possible sur serveur |

**Epreuve 2 - 2ème Partie**

***Mise en oeuvre du projet***

**Expliquez les principales étapes de la mise en oeuvre de votre projet :**

|  |
| --- |
| - Reprise du pipeline entraîné (prétraitement + modèle) - Création d’un script app\_gradio.py intégrant le modèle et les étapes de preprocessing - Création d’une fonction predict\_offer() prenant les inputs du formulaire - Test avec différentes entrées utilisateurs pour valider la robustesse - Interface ergonomique, affichant les prédictions de manière lisible |

**Epreuve 2 - 3ème Partie**

***Bilan du projet***

**Quel bilan tirez-vous de votre projet ? :**

|  |
| --- |
| Le modèle final (XGBoost) atteint un score d’accuracy supérieur à 87 % sur les classes principales. La vectorisation des titres via TF-IDF et la combinaison avec des données tabulaires offre une bonne complémentarité. |

**Epreuve 2 - 3ème Partie**

***Axes d’améliorations***

**Si vous aviez 1 mois pour finaliser votre projet, quelles sont les fonctionnalités que vous ajouteriez ? :**

|  |
| --- |
| - Déploiement sur un serveur distant avec gestion des dépendances (Docker ou Heroku) - Ajout d’une base de données pour stocker les prédictions - Ajout d’un système de feedback utilisateur pour affiner le modèle |

**Epreuve 2 - Conclusion**

|  |
| --- |
| Cette formation m’a permis de consolider mes compétences en traitement de données, modélisation IA et mise en production. J’ai appris à structurer un projet IA de bout en bout, avec une forte exigence de rigueur dans le code et la documentation. Je suis désormais capable de proposer une solution IA complète, du nettoyage de données jusqu’à la mise à disposition via une application. Je n’ai aucun regret, seulement l’envie d’aller plus loin sur les sujets de MLOps et d’industrialisation. |