

Auto-évaluation

Classifiez automatiquement des informations



Un dernier doute avant l'envoi de vos livrables ?

Pour vérifier la qualité de votre travail :

- cochez les cases ci-dessous : elles indiquent que vous avez bien pris en compte chaque indicateur de réussite ;
- renseignez, si besoin, la colonne "Notes" avec des commentaires sur vos livrables / vos étapes. Ils seront des points de discussion avec votre mentor pendant votre session de bilan / soutenance.

Quand toutes les cases de ce document seront cochées, vous pourrez déposer vos livrables sur la plateforme.

Bonne réussite !

Compétences	Livrables	Indicateurs de réussite de l'activité	Notes
Configurer l'environnement de travail nécessaire à l'exploitation des données	<ul style="list-style-type: none"> Fichier pyprojectct.toml 	<input type="checkbox"/> J'ai utilisé uv ou Poetry pour gérer les dépendances et créer l'environnement virtuel Python hébergeant le projet.	
	<ul style="list-style-type: none"> Notebook avec les packages pertinents importés 	<input type="checkbox"/> J'ai importé les packages requis dans un notebook au sein de l'environnement virtuel que j'ai créé.	
Mettre en place un processus de nettoyage afin d'améliorer la qualité de données	<ul style="list-style-type: none"> Notebook : <ul style="list-style-type: none"> DataFrame central issu d'un rapprochement des 3 fichiers initiaux 	<input type="checkbox"/> J'ai nettoyé les colonnes du jeu de données pour faciliter les analyses suivantes. <input type="checkbox"/> J'ai identifié les colonnes nécessaires pour réaliser une jointure adaptée entre les 3 jeux de données.	
	<ul style="list-style-type: none"> Notebook : <ul style="list-style-type: none"> Analyse descriptive des employés 	<input type="checkbox"/> J'ai dégagé des statistiques clés sur les employés, en distinguant ceux qui ont quitté l'entreprise des autres. <input type="checkbox"/> J'ai créé des graphiques pertinents pour illustrer les différences entre les deux populations principales d'employés.	
Préparer et transformer des données afin de les adapter au modèle d'apprentissage	<ul style="list-style-type: none"> Notebook <ul style="list-style-type: none"> DataFrame de features (X) et Pandas Series contenant la cible (y) 	<input type="checkbox"/> J'ai appliqué des techniques d'encodage adaptées aux features qualitatives. <input type="checkbox"/> J'ai créé quelques features à partir des données existantes.	

		<input type="checkbox"/> J'ai utilisé une matrice de corrélation de Pearson pour identifier et éventuellement traiter certaines redondances dans les features. <input type="checkbox"/> J'ai réalisé un choix adapté de scaling des features.	
Entraîner un modèle d'apprentissage	<ul style="list-style-type: none"> • Notebook <ul style="list-style-type: none"> ◦ Modèles entraînés 	<input type="checkbox"/> J'ai entraîné plusieurs modèles, dont un étalon, un linéaire et un non-linéaire.	
	<ul style="list-style-type: none"> • Notebook <ul style="list-style-type: none"> ◦ Métriques d'évaluation 	<input type="checkbox"/> J'ai calculé sur plusieurs jeux d'apprentissages et de train, plusieurs métriques de performance adaptées à ma problématique.	
	<ul style="list-style-type: none"> • Notebook <ul style="list-style-type: none"> ◦ Approche de modélisation 	<input type="checkbox"/> J'ai adapté ma méthodologie pour prendre en compte le déséquilibre des classes. <input type="checkbox"/> J'ai déterminé un seuil de probabilité adapté pour permettre au modèle de classer des individus. <input type="checkbox"/> J'ai optimisé les hyperparamètres de l'algorithme pour en améliorer la performance.	
Évaluer le modèle d'apprentissage	<ul style="list-style-type: none"> • Notebook <ul style="list-style-type: none"> ◦ Classement de l'importance des features 	<input type="checkbox"/> J'ai utilisé une méthode de feature importance globale pour classer en termes de pertinence les features en entrée du modèle. <input type="checkbox"/> J'ai calculé les valeurs de Shapley en utilisant le package SHAP.	

		<input type="checkbox"/> J'ai utilisé le graphique Beeswarm pour générer un autre classement de la feature importance.	
	<ul style="list-style-type: none"> • Notebook <ul style="list-style-type: none"> ○ Explications potentielles des causes de démission 	<input type="checkbox"/> J'ai utilisé des graphiques adaptés pour quantifier plus finement l'impact des features sur le modèle, y compris sur des individus des deux classes.	
Communiquer et présenter les résultats	<ul style="list-style-type: none"> • Présentation du projet 	<input type="checkbox"/> J'ai rappelé de manière concise le contexte du projet et le contenu de la donnée de départ. <input type="checkbox"/> J'ai synthétisé les analyses descriptives en choisissant les plus pertinentes par rapport à la problématique métier. <input type="checkbox"/> J'ai expliqué de manière claire les choix méthodologiques de modélisation et les résultats. <input type="checkbox"/> J'ai restitué de manière claire et compréhensible pour une audience non-technique, les causes potentielles de démission au sein de l'entreprise.	