

Ce document vous explique ce qui est attendu de vous durant la réalisation de votre projet. Vous aurez trois rendus principaux :

- **rendu 1** : introduire, décrire, visualiser et manipuler votre jeu de données ;
- **rendu 2** : modélisation et résolution de la problématique à l'aide de techniques de Machine Learning ;
- **rapport final** : constitué des deux premiers rendus, il concrétise votre vision et votre travail face à votre projet en intégrant une conclusion et une ouverture.

Cette méthodologie a donc vocation à vous conseiller dans la réalisation des rendus, mais ne représente pas des consignes précises qu'il vous faut suivre. Vous êtes maîtres de votre travail, de vos analyses, ainsi que du format que prendront vos différents rapports.

Rendu 1 : rapport d'exploration, de data visualisation et de pre-processing des données

Introduction au projet

Contexte

- Contexte d'insertion du projet dans votre métier.
- Du point de vue technique.
- Du point de vue économique.
- Du point de vue scientifique.

Objectifs

- Quels sont les principaux objectifs à atteindre ? Décrivez en quelques lignes.
- Pour chacun des membres du groupe, préciser le niveau d'expertise autour de la problématique adressée ?
- Êtes-vous entré en contact avec des experts métiers pour affiner la problématique et les modèles sous-jacents ? Si oui, détaillez l'apport de ces interactions.
- (Avez-vous connaissance d'un projet similaire au sein de votre entreprise, ou bien dans votre entourage ? Quel est son état d'avancement ? En quoi vous a-t-il aidé dans la réalisation de votre projet ? En quoi votre projet contribue-t-il à l'améliorer ?)

Compréhension et manipulation des données

Cadre

- Quel(s) jeu(x) de donnée(s) avez vous utilisé pour atteindre les objectifs de votre projet ?
- Ces données sont-elles disponibles librement ? Dans le cas contraire, qui est le propriétaire de la donnée ?
- Décrivez la volumétrie de votre jeu de données ?

Pertinence

- Quelles variables vous semblent les plus pertinentes au regard de vos objectifs ?
- Quelle est la variable cible ?
- Quelles particularités de votre jeu de données pouvez-vous mettre en avant ?
- Etes-vous limités par certaines de vos données ?

Pre-processing et feature engineering

- Avez-vous eu à nettoyer et à traiter les données ? Si oui, décrivez votre processus de traitement.
- Avez-vous dû procéder à des transformations de vos données de type normalisation/standardisation ? Si oui, pourquoi ?
- Envisagez-vous des techniques de réduction de dimension dans la partie de modélisation ? Si oui, pourquoi ?

Visualisations et Statistiques

- Avez-vous identifié des relations entre différentes variables ? Entre variables explicatives ? et entre vos variables explicatives et la/les cible(s) ?
- Décrivez la distribution de ces données, répartition, outliers.. (pré/post traitement éventuellement)
- Présentez les analyses statistiques utilisées permettant de confirmer les informations présentes sur les graphiques.
- Tirer des conclusions des éléments relevés ci-dessus permettant de se projeter dans la partie de modélisation

Modalités d'évaluation :

Mise en situation professionnelle reconstituée : à partir d'un jeu de données d'entreprise, le/la candidate(e) devra mettre en œuvre divers prétraitements et augmentation de données afin de rendre ces dernières exploitables par les techniques d'apprentissage automatisées.

Rendu 2 : rapport de modélisation

Étapes de réalisation du projet

Classification du problème

- À quel type de problème de machine learning votre projet s'apparente-t-il ? (classification, régression, clustering, etc)
- À quelle tâche de machine learning votre projet s'apparente-t-il ? (détection de fraude, reconnaissance faciale, analyse de sentiment, etc) ?
- Quelle est la métrique de performance principale utilisée pour comparer vos modèles ? Pourquoi celle-ci ?
- Avez-vous utilisé d'autres métriques de performances qualitative ou quantitative ? Si oui, détaillez le.

Choix du modèle et optimisation

- Quels algorithmes avez-vous essayés ?
- Décrivez celui / ceux que vous avez retenu et pourquoi ?
- Avez-vous utilisé des techniques d'optimisation de paramètres de type Grid Search et Validation Croisée ?
- Avez-vous testé des modèles avancés ? Bagging, Boosting, Deep Learning... Pourquoi ?

Interprétation des résultats

- Avez-vous analysé les erreurs de votre modèle ?
- Cela a-t-il contribué à son amélioration ? Si oui, décrivez.
- Avez-vous utilisé des techniques d'interprétabilité de type SHAP, LIME, Skater... (Grad-CAM pour le Deep Learning..)
- Qu'est ce qui a (ou non) engendré une amélioration significative de vos performances ?

Modalités d'évaluation :

Mise en situation professionnelle : en se basant sur une solution proposée, le/la candidat(e) devra réaliser un rapport de synthèse incluant : l'explication des choix de solutions IA implémentés, l'interprétation des résultats, l'évaluation de la fiabilité des algorithmes et une proposition d'optimisation.

Rapport final :

Conclusion tirées

Difficultés rencontrées lors du projet

- Quel a été le principal verrou scientifique rencontré lors de ce projet ?
- Pour chacun des points suivants, si vous avez rencontré des difficultés, détaillez en quoi elles vous ont ralenti dans la mise en place de votre projet.
- Prévisionnel : tâches qui ont pris plus de temps que prévu, etc.
- Jeux de données : acquisition, volumétrie, traitement, agrégation, etc.
- Compétences techniques / théoriques : timing d'acquisition des compétences, compétence non proposée en formation, etc..
- Pertinence : de l'approche, du modèle, des données, etc..
- IT : puissance de stockage, puissance computationnelle, etc.
- Autres

Bilan

- Détaillez quelle a été votre contribution principale dans l'atteinte des objectifs du projet.
- Avez-vous modifié le modèle depuis la dernière itération ? Si oui, détaillez.
- Présentez les résultats obtenus et comparez-les au benchmark
- Pour chacun des objectifs du projet, détaillez en quoi ils ont été atteints ou non.
- S'ils ont été atteints, dans quel(s) process(es) métier(s) votre modèle peut-il s'inscrire ? Détaillez.

Suite du projet

- Quelles pistes d'amélioration suggérez-vous pour augmenter les performances de votre modèle ?
- En quoi votre projet a-t-il contribué à un accroissement de la connaissance scientifique ?

Bibliographie

- Sur quels éléments bibliographiques (articles de recherches, blog, livres, etc) vous êtes vous appuyé pour réaliser votre projet ?

Annexes

- Diagramme de Gantt.
- Description des fichiers de code.