



Année universitaire	2025-2026		
Département	Informatique	Année	M1
Matière	Apprentissage et Analyse de données		
Enseignant	Haytham Elghazel & Ichrak Ennaceur		
Durée	3h		
Évaluation	Il faut un code bien organisé ainsi que votre prédiction sur la base scoring Rendu uniquement sur Moodle. Aucun Rendu ne sera accepté par Mail !!! Tout autre format ou nom incorrect pour le fichier rendu ne sera pas accepté pour la notation !!!		

L'objectif dans ce travail est de construire un bon classifieur sur un jeu de données de *Surgery* du fichier "**Surgery_train.csv**". On cherche si un patient aura des complications pendant une opération chirurgicale ou non (**complication**). Votre classifieur a pour objectif d'optimiser une fonction coût donnée par la formule suivante :

$$0.5 * \text{Rappel}(\text{classe } 0) + 0.5 * \text{Rappel}(\text{classe } 1)$$

Une fois que vous avez choisi votre meilleure stratégie d'apprentissage sur votre jeu de données, vous devez fournir une prédiction en utilisant la base de scoring du fichier "**Surgery_scoring.csv**".

Voici quelques informations sur les variables explicatives du jeu de données :

- **bmi** : indice de masse corporelle.
- **age** : Age du patient.
- **asa_status** : Statut physique selon l'American Society of Anesthesiologists (ASA).
- **baseline_cancer** : Antécédent de cancer.
- **baseline_charlson** : Indice de comorbidité de Charlson (mesure du risque basé sur les comorbidités).
- **baseline_cvd** : Antécédent de maladie cardiovasculaire.
- **baseline_dementia** : Antécédent de démence.
- **baseline_diabetes** : Antécédent de diabète.
- **baseline_digestive** : Antécédent de maladie digestive.
- **baseline_osteoart** : Antécédent de maladie ostéo-articulaire (ex : arthrose, arthrite).
- **baseline_psych** : Antécédent de trouble psychiatrique.
- **baseline_pulmonary** : Antécédent de maladie pulmonaire (ex : asthme, BPCO).
- **ahrq_ccs** : Catégorie de procédure selon le système AHRQ-CCS (Clinical Classifications Software). Elle permet de regrouper les procédures médicales pour analyse statistique des complications et de la mortalité.
- **ccsComplicationRate** : Taux global de complications intra-hospitalières pour chaque catégorie de procédure AHRQ-CCS.
- **ccsMort30Rate** : Taux global de mortalité à 30 jours pour chaque catégorie de procédure AHRQ-CCS.
- **complication_rsi** : Risk Stratification Index – Indice de risque de complications intra-hospitalières.
- **hour** : Heure de l'intervention / admission.
- **dow** : Jour de la semaine.
- **genre** : Sexe du patient.
- **month** : Mois de l'intervention.
- **moonphase** : Phase de la lune.

- 0 = nouvelle lune
- 1 = premier quartier
- 2 = pleine lune
- 3 = dernier quartier
- **mort30** : Mortalité à 30 jours.
- **mortality_rsi** : Risk Stratification Index – Indice de risque de mortalité à 30 jours. Il permet d'estimer le risque de décès post-opératoire sur la base des caractéristiques du patient et de la procédure.
- **race** : Race / groupe ethnique du patient.
 - 0 = Caucasian
 - 1 = Afro-American
 - 2 = Autre
- **temporal_risk_score** : Score synthétique basé sur le contexte temporel et organisationnel de l'intervention (jour de la semaine, heure, mois).

Votre rendu final attendu est :

1. le code d'apprentissage (sous forme de notebook python).
2. le fichier **CSV** comprenant deux colonnes pour chaque instance du fichier de scoring à savoir son ID et sa classe prédite sous la forme (PatientID, classe).

Code pour création du fichier CSV : `X_scoring.to_csv('Surgey_scoring_predictions_Nom_Prenom.csv', index=False, columns=['PatientID', 'classe'])`

3. **Attention :** le fichier de prédiction doit être nommé exactement comme suit : **Surgey_scoring_predictions_Nom_Prenom.csv**. Remplacez Nom_Prenom par votre nom et prénom.