**Program of the booklet**

**1er jour :**

9h00 : 9h15🡪 Mot de bienvenue

9h15 :10h15🡪 Présentation des participants

Pause-café : 10:15 - 10:30

10h30 : 11h00 🡪Introduction générale

* Introduction à R et RStudio (présentation et installation)

11h :11h15🡪Bases du langage R :

* Variables
* Types de données
* Opérateurs

11h15 :11h30 🡪 Fonctions et packages :

* Fonctions
* Packages

11h30 :13h🡪Manipulations de données avec R :

* Importation des données
* Exploration et manipulation des données (principe de slicing…)
* Introduction à dplyr pour le traitement des données

13h :14h🡪Diner d’affaires

14h :16h🡪Nettoyage et transformation des données :

* Traitement des valeurs manquantes
* Suppression des doublons
* Traitement des outliers
* Vérification des types de données
* Vérification de la cohérence des données
* Autres vérifications (Fusionner ou fractionner des colonnes…)

16h :17h🡪Réseautage

**2e jour** :

9h30 :10h15🡪Statistiques descriptives :

* Indicateurs de tendance centrale
* Indicateur de variabilité
* Quantile
* Table de contingence
* Autres indicateurs (coefficient/matrice de corrélation, matrice de variance covariance

Pause-café : 10:15 - 10:30

10h30 :11h30🡪Statistiques inférentielles :

* Test chi2
* Test de student
* Analyse de la variance
* Test de Fischer

11h :12h🡪Visualisation :

* Scatter plot
* Histogramme
* Bar plot
* Boxplot
* Map visualisation
* Line chart

12h :13h🡪Introduction au machine learning et techniques de machine learning

* Introduction générale
* Techniques de machine learning
* Notions d’underfitting et d’overfitting

13h :14h Diner d’affaires

14h :16h🡪Préparation des données :

* Standardisation et normalisation
* Selection des variables pertinentes(analyse de corrélation,RFE,….)
* Gestion des variables catégorielles
* Reduction de données
* Gestion des données déséquilibrées

16h :17h Réseautage

**3e jour** :

9h30 :10h30🡪Modélisation et recherche des hyperparamètres optimaux :

* Notion de train, test and validation set
* Notion de cross validation
* Formation du modèle
* Hyperparameter tuning

Pause-café : 10:15 - 10:30

10h30 : 11h 🡪 Modélisation et recherche des hyperparamètres optimaux (suite)

11h :11h45🡪Evaluation des modèles et sélection du modèle optimal :

* Métriques d’évaluation
* Critères de sélection d’un modèle optimal

11h45 :12h15🡪Utilisation des modèles pour les prédictions

Deploiement du modèle