

Table des matières

1 — Prérequis techniques	3
1.1. Bibliothèques utiles	3
1.2. Format de restitution	4
2 — Sujet	5
2.1. Les données	6
2.1.1. Diagnostic	6
2.2. Aides au code	7
2.2.1. Importer les données	7
3 — Evaluation	8

1 — Prérequis techniques

En accord avec les professeurs des deux universités, les deux langages de programmation autorisés pour résoudre la problématique proposée sont Python et R. Toute autre technologie utilisée ne fera donc pas l'objet d'une évaluation.

1.1. Bibliothèques utiles



Packages	Utilisation	Documentation
Numpy	Calculs scientifiques et matriciels	https://numpy.org/doc/stable/
Pandas	Charger et manipuler des données	https://pandas.pydata.org/docs/
Matplotlib	Visualisation	https://matplotlib.org/stable/users/index.html
Scikit Learn	Machine learning	https://scikit-learn.org/stable/user_guide.html
Tensorflow	Machine learning et Deep learning	https://www.tensorflow.org/?hl=fr
Pytorch	Machine learning et Deep learning	https://pytorch.org/



CRAN

Packages	Utilisation	Documentation
Matrix	Calculs matriciels	https://cran.r-project.org/web/packages/Matrix/index.html
Plyr	Manipuler des données	https://cran.r-project.org/web/packages/plyr/index.html
DataExplorer	Exploration et traitement de données	https://cran.r-project.org/web/packages/DataExplorer/index.html
Machine-learning in R	Machine learning	https://www.datacamp.com/tutorial/machine-learning-in-r
Ggplot2	Visualisation	https://ggplot2.tidyverse.org/
Nnet	Deep Learning	https://cran.r-project.org/web/packages/ggplot2/index.html

Ces listes ne sont en aucun cas exhaustives. D'autres librairies peuvent être utilisées si nécessaire. De plus, il ne s'agit pas de modules obligatoires. Certains peuvent ne pas être utiles selon les choix de résolution des cas d'usage.

1.2. Format de restitution

Le choix concernant les supports de restitution des travaux est laissé libre aux participants. Il serait tout de même préférable de réaliser deux livrables distincts : un document de synthèse des analyses et des réflexions menées ainsi que le script de code réalisé. Pour faciliter la compréhension de ce dernier, on peut utiliser des fichiers au format Notebook (type Jupyter Notebook en python) qui permettent de créer et de partager des documents interactifs contenant du code dynamique et exécutable, des visualisations de contenus, des textes de documentation et des équations.

Pour plus de précisions, se référer aux consignes établies dans le règlement du challenge.

2 — Sujet

Comme présenté lors de la phase d'introduction du challenge, le premier cas d'usage présenté porte sur l'optimisation des survols du réseau pour diagnostic.

Chaque année Enedis survole 1/3 du réseau aérien moyenne tension breton afin de visiter la totalité du réseau sur 3 ans. Ce premier survol généralement fait par hélicoptère sert à faire de la prise d'images et détecter des anomalies structurelles.

La seconde étape pour le service responsable de la rénovation programmée est de faire un dépouillement, c'est-à-dire de définir les tronçons à visiter en drone pour diagnostic (max 150 km de ligne). Pour ce faire il s'appuie sur les anomalies relevées mais aussi sur les incidents survenus l'année passée, le plan aléa climatique et l'état du réseau. Les incidents représentent toutes les coupures ayant une cause réseau. Le plan aléa climatique permet de définir les sections de ligne sensible au climat comme celles qui sont par exemple très exposées à la contrainte vent. L'état du réseau lui décrit la ligne électrique (tronçon) comme par exemple sa longueur ou sa date de mise en exploitation. Suite à l'obtention des diagnostics, les maintenances qui s'imposent sont prévues pour l'année à venir.

Pour réaliser cet exercice toutes les données anomalies, incidents, plan aléa climatique et état du réseau sont récupérées sur l'année N-1 et on observe les maintenances réalisées sur l'année N pour s'assurer que les survols en drone ont été correctement priorisés.

On souhaite donc déterminer les tronçons à prioriser pour remplacer le dépouillement manuel.

2.1. Les données

2.1.2. Etat du réseau

L'intérêt de l'entreprise est de visiter en priorité les tronçons comportant un maximum de faiblesses, aussi bien en termes d'état, d'incidents sur le réseau, d'anomalies détectées ou de sensibilité aux aléas climatiques. La base de données suivante, [Diagnostic.xlsx](#) contient les informations clés permettant de juger de la faiblesse d'un tronçon. Toutes les données ont été générées sur l'année 2022 sauf l'immobilisation RP sur l'année 2023.

Attention :

- S'il des données sont manquantes c'est qu'aucun événement ne s'est produit.

Format :

Colonne	Description courte
ID_t	Identifiant du tronçon
Nb_of_incident	Nombre d'incident sur le départ
Electrical_length	Longueur électrique du tronçon (mètre)
Service_date	Date de mise en exploitation du tronçon
Length_climate_hazard_plan	Longueur en plan aléa climatique sur le départ
Length_fragile_section	Longueur section fragile sur le départ
Nb_of_anomaly	Nombre d'anomalie sur le départ
Year_helicopter_flight	Année du dernier vol sur le départ
Last treatment PR immobilized	Année de la dernière maintenance sur le départ

2.2. Aides au code

2.2.1. Importer les données



```
# modules nécessaires
import pandas as pd

# lecture des données au format xlsx
df = pd.read_excel('Diagnostic.xlsx', delimiter=';', encoding='latin-1')

# les données sont ensuite exploitables sous la forme d'un dataframe
classique
```



```
# lecture du fichier
df <- read.csv("Diagnostic.xlsx", encoding="latin-1", sep=";")
```

3 — Evaluation

Vous possédez 2 bases données. Une première **Diagnostic_train.xlsx** et une deuxième **Diagnostic_test_etu.xlsx** sans la sortie. La première vous servira à entraîner vos modèles et la deuxième servira aux évaluateurs à mesurer l'erreur quadratique moyenne de vos modèles (MSE) ainsi que la longueur maximale de lignes (25 km max).

Pour évaluer vos modèles les évaluateurs auront besoin d'un fichier Excel de votre part contenant les colonnes : **ID_t** pour l'identifiant du tronçon et **y_pred** contenant vos prédictions de type Int64 (1 s'il faut prévoir un diagnostic et 0 sinon). Vous nommerez le fichier : **Diagnostic_pred_etu.xlsx**.

Bonus :

L'idéal serait que la somme des longueurs des tronçons à diagnostiquer ne dépasse pas les 25 km pour la base de test.