# Rapport de modélisation

## Proposition de plan

Légende :

* Violet : Alex
* Or : Omar

1. Modélisation
   1. Méthodologie
      1. Phases
         1. Phase A : « 9df »
            * Rappeler création des 9 sous-ensembles (§ 2.2.2)
            * Illustrer préparation des données (data\_prep\_pipeline.png)
         2. Phase B : « 1df »
            * Décrire problèmes rencontrés dans phase A (suivi fastidieux, comparabilité problématique, etc.)
            * Présenter pipeline simplifié de préparation des données (+ illustrer ?)
         3. Phase C : choix de métrique et optimisation
            * Expliquer pourquoi nous avons choisi precision\_1 = 0,83 comme métrique
            * Résumer nos tentatives d’amélioration des scores (mise à l’échelle, rééquilibrage, etc.)
            * Faire le bilan de ces essais (globalement mitigé, varie selon les algorithmes et les paramètres)
      2. Approches
         1. Classification : algorithmes classiques
            * LogisticRegression
            * DecisionTreeClassifier
            * RandomForestClassifier
            * KNeighborsClassifier
            * SVC
         2. Classification : algorithmes des séries temporelles
            * K-Nearest Neighbors avec Dynamic Time Warping (KNN-DTW)
            * Time Series Forest (TSF)
            * RandOm Convolutional KErnel Transform (ROCKET)
            * Prétraitement complémentaire des données : tri, découpage, conversion
         3. Méthodes d’ensemble
            * VotingClassifier (hard)
            * StackingClassifier
            * XGBoost
         4. Apprentissage profond
            * Réseaux de neurones profonds
         5. ~~Bonus : régression linéaire~~
   2. Résultats
      1. Décrire scores.xlsx (+ illustrer ?)
   3. Discussion
      1. Interprétation des résultats
         * + Les rapporter à leurs jeux de données respectifs (9df vs 1df)
           + Évoquer les effets des optimisations
           + Indiquer que nous avons décidé de retenir 1 algorithme par famille d’approches (RandomForestClassifier, TimeSeriesForestClassifier, VotingClassifier (hard), réseaux de neurones profonds)
           + Pourquoi ces algorithmes semblent-ils mieux adaptés à notre jeu de données que les autres ?
      2. Remise en contexte
         * + Quels sont les algorithmes d’apprentissage automatique / profond qui sont réellement utilisés dans le domaine de la prévision météorologique ?
           + Sont-ils les mêmes que ceux que nous avons retenus ?
           + Quel rôle joue l’interprétabilité des algorithmes / modèles dans le choix de ces derniers dans le domaine de la prévision météorologique ?
   4. Perspectives
      1. Preprocessing
         * + Gérer les NaN plus finement, notamment *via* l’interpolation temporelle
           + Supprimer des données, voire des variables
      2. Feature engineering
      3. Optimisations
      4. Régression linéaire