## 1.3. Discussion

### 1.3.1. Interprétation des résultats

Comme indiqué dans la section 4.2 (Résultats), les meilleurs scores de precision\_1 viennent des **méthodes d’ensemble** (Voting, Stacking, XGBoost) et de la **classification des séries temporelles** (ROCKET).

Concernant les méthodes d’ensemble en particulier, et comme évoqué dans la section 4.1.2.3 (Méthodes d’ensemble), notre implémentation du Voting réunit et fait voter les quatre modèles optimisés suivants :

1. LogReg
2. DT
3. RF
4. KNN

Notre implémentation du Stacking, qui affiche également de belles performances, est construite sur ces mêmes modèles de base, suivie de la désignation de LogReg comme estimateur final.

**Le pouvoir de ces approches** résident en leur capacité à combiner des modèles très différents : elles sont en mesure de tirer profit des forces individuelles de ces derniers, tout en maîtrisant leurs faiblesses respectives. Cette diversité, que nous avons essayé de respecter au mieux dans nos implémentations (en choisissant des modèles aussi variés que possible), est un atout face notre jeu de données. En effet, la complexité, le déséquilibre, les dimensions et l’hétérogénéité de ce dernier mettent à mal les modèles classiques, qui ne parviennent pas à en capturer les caractéristiques plurielles et la structure complexe sous-jacente.[[1]](#endnote-1)

La bonne place de ROCKET dans le classement peut, elle aussi, être attribuée à la capacité de cette méthode à appréhender des motifs très variés, grâce à la forte diversité des paramètres de ses milliers de noyaux de convolution.[[2]](#endnote-2)

Cependant, malgré leurs bons scores, ces modèles comportent aussi des **inconvénients.** Les deux principaux sont les suivants :

1. **Ils sont coûteux en ressources de calcul.[[3]](#endnote-3)** À titre d’exemple, notre implémentation du Voting revient à faire quatre modélisations, ce qui représente une opération conséquente compte tenu des dimensions de notre jeu de données. Le même argument s’applique à ROCKET : la transformation des données et l’entraînement du modèle de classification linéaire ont nécessité beaucoup de mémoire vive et un temps d’exécution très long.
2. **Leur complexité se traduit par une interprétabilité plus faible.[[4]](#endnote-4)** Dans le domaine de l’apprentissage automatique, la question de l’interprétabilité des modèles est loin d’être anodine. En fonction du secteur d’activité, des parties prenantes et d’autres éléments de contexte, elle peut même être centrale et contraindre le choix de modèle (souvent au détriment de la performance, laquelle augmente généralement avec la sophistication d’un modèle). Le plus souvent, le besoin ou le désir de développer la confiance des usagers dans un modèle et ses prédictions incite à en préférer un avec une bonne interprétabilité, quitte à faire des concessions sur la performance.[[5]](#endnote-5) L’environnement réglementaire influe également sur les parties prenantes : le RGPD donne, aux personnes dont les données sont traitées, le droit à une explication des décisions algorithmiques et le droit d’être informées.[[6]](#endnote-6) Enfin, le fait de proposer des interprétations et des explications des algorithmes souvent perçus ou caractérisés comme des « boîtes noires » ont, en soi, une valeur sociale et éthique.[[7]](#endnote-7)

Il n’est donc pas évident de préférer les méthodes d’ensemble et ROCKET aux autres approches que nous avons également testées, notamment celles basées sur les **arbres de décision,** car ces dernières proposent des scores qui restent corrects (même s’ils sont en-dessous du seuil de 83% que nous avons fixé pour precision\_1), tout en étant plus facilement interprétables.

Afin de nous assurer que nos choix finaux seront faits à partir des meilleurs scores possibles, nous avons effectué diverses **optimisations** (mise à l’échelle, rééquilibrage, optimisation du seuil). Pour résumer les observations que nous avons présentées dans la section 4.2 (Résultats), ces optimisations fonctionnent très bien dans la plupart des cas, avec des **gains pouvant aller jusqu’à 15%.** En revanche, elles ont l’effet contraire sur les algorithmes de **classification des séries temporelles,** avec des **pertes pouvant aller jusqu’à 66%.** Cela est peut-être dû à notre emploi d’**approches naïves** au sein des étapes suivantes de prétraitement des données :

1. **Mise à l’échelle.** Il existe de nombreuses méthodes de normalisation des séries temporelles, chacune adaptée à des problématiques différentes.[[8]](#endnote-8) Par manque de temps et d’expérience, nous nous sommes limités aux méthodes proposées par scikit-learn (StandardScaler, MinMaxScaler). Par ailleurs, bien que la normalisation soit généralement nécessaire dans l’étude des séries temporelles, elle peut aussi détruire des informations et contribuer à la dégradation des analyses.[[9]](#endnote-9)
2. **Rééquilibrage.** Notre approche ne tient pas compte de la difficulté inhérente à la classification des séries temporelles déséquilibrées, une thématique de recherche qui se révèle assez riche et complexe.[[10]](#endnote-10) En effet, la dimensionnalité élevée et la forte corrélation inter-variable qui caractérisent ces ensembles de données ne sont pas prises en compte par les méthodes classiques proposées par scikit-learn (SMOTE, RandomUnderSampler).[[11]](#endnote-11)

Compte tenu de tous les éléments exposés ci-dessus, et afin de néanmoins pouvoir effectuer des comparaisons, nous avons finalement décidé de retenir **un algorithme pour chacune des quatre familles d’approches** que nous avons mobilisées dans le cadre de ce projet :

1. Classification (algorithmes classiques) : **RF**
2. Classification (algorithmes des séries temporelles) : **ROCKET**
3. Méthodes d’ensemble : **Soft Voting**
4. Réseaux de neurones : **DNN**

### 1.3.2. Remise en contexte

La ***numerical weather prediction* / NWP** (prévision numérique du temps / PNT en français) utilise des modèles mathématiques de l’atmosphère et des océans pour prédire le temps en fonction des conditions météorologiques actuelles.[[12]](#endnote-12) Son histoire remonte à 1922, mais ce n’est qu’à partir de 1950, avec l’apparition des premiers ordinateurs, qu’elle a réellement commencé à se développer.[[13]](#endnote-13)

Bien que la NWP reste encore aujourd’hui l’approche standard en prédiction météorologique, les météorologues s’intéressent de plus en plus à l’apprentissage machine / profond, certains ayant commencé à intégrer ces approches dans leurs flux de travail.[[14]](#endnote-14) D’autres spécialistes restent sceptiques et réticents, reprochant à l’apprentissage automatique / profond un manque d’explicabilité (notamment dans le cas des réseaux de neurones profonds), ainsi que l’absence de prise en compte des contraintes physiques.[[15]](#endnote-15) Toutefois, cela n’empêche pas la croissance de l’appétit des météorologues pour les deux approches,[[16]](#endnote-16) avec une accélération depuis février 2022.14

Parmi les deux, **l’apprentissage profond** est de plus en plus plébiscité (certains travaux estiment que l’apprentissage automatique est moins adapté aux données géospatiales[[17]](#endnote-17)). En novembre 2023, Google DeepMind a publié un article dans *Science* qui compare les performances de GraphCast, son modèle de prédiction météorologique basé sur les *graph neural networks* (réseaux de neurones graphiques), à celles de l’Integrated Forecasting System (IFS), le modèle NWP du European Centre for Medium-Range Weather Forecasts (ECMWF) qui était considéré jusqu’à présent comme l’étalon-or de la discipline.[[18]](#endnote-18) GraphCast a distancé de loin l’IFS en matière d’*accuracy* et de vitesse. L’ECMWF a reconnu que les modèles d’apprentissage profond sont devenus tellement performants qu’ils sont désormais incontournables en météorologie, et l’organisme a même déjà commencé à utiliser GraphCast.[[19]](#endnote-19) Ils expérimentent également avec d’autres modèles de pointe, notamment Pangu-Weather de Huawei[[20]](#endnote-20) (basé sur les *3D neural networks*) et FourCastNet de NVIDIA[[21]](#endnote-21) (basé sur les *Fourier-based neural networks*).14

Nous avons également tenté d’appliquer l’apprentissage profond à notre problématique de projet avec notre modèle DNN, mais les résultats ne sont pas à la hauteur de nos espérances. Cela ne veut pas dire que la méthode n’est pas adaptée (d’autres chercheurs ayant implémenté des réseaux de neurones simples ont formulé cette conclusion hâtivement15), mais qu’il faut trouver la bonne architecture et les bons paramètres, ce qui correspond au véritable défi à la maîtrise du domaine complexe qu’est l’apprentissage profond.

1. <https://link.springer.com/article/10.1007/s11704-019-8208-z> [↑](#endnote-ref-1)
2. <https://pub.towardsai.net/rocket-fast-and-accurate-time-series-classification-f54923ad0ac9> [↑](#endnote-ref-2)
3. <https://towardsdatascience.com/combine-your-machine-learning-models-with-voting-fa1b42790d84> [↑](#endnote-ref-3)
4. <https://towardsdatascience.com/use-voting-classifier-to-improve-the-performance-of-your-ml-model-805345f9de0e> [↑](#endnote-ref-4)
5. <https://wires.onlinelibrary.wiley.com/doi/full/10.1002/widm.1493> [↑](#endnote-ref-5)
6. <https://link.springer.com/book/10.1007/978-3-319-57959-7> [↑](#endnote-ref-6)
7. <https://heinonline.org/HOL/LandingPage?handle=hein.journals/hjlt31&div=29&id=&page=> [↑](#endnote-ref-7)
8. <https://www.sciencedirect.com/science/article/abs/pii/S2214579623000400> [↑](#endnote-ref-8)
9. <https://content.iospress.com/articles/journal-of-intelligent-and-fuzzy-systems/ifs171393> [↑](#endnote-ref-9)
10. <https://ieeexplore.ieee.org/abstract/document/6468038> [↑](#endnote-ref-10)
11. <https://www.sciencedirect.com/science/article/pii/S0950705122003586> [↑](#endnote-ref-11)
12. <https://en.wikipedia.org/wiki/Numerical_weather_prediction> [↑](#endnote-ref-12)
13. <https://www.sciencedirect.com/science/article/abs/pii/S0021999107000952> [↑](#endnote-ref-13)
14. <https://www.ecmwf.int/en/about/media-centre/science-blog/2023/rise-machine-learning-weather-forecasting> [↑](#endnote-ref-14)
15. <https://royalsocietypublishing.org/doi/full/10.1098/rsta.2020.0097> [↑](#endnote-ref-15)
16. <https://www.ecmwf.int/en/about/media-centre/news/2023/how-ai-models-are-transforming-weather-forecasting-showcase-data> [↑](#endnote-ref-16)
17. <https://www.nature.com/articles/s41586-019-0912-1> [↑](#endnote-ref-17)
18. <https://www.science.org/doi/10.1126/science.adi2336> [↑](#endnote-ref-18)
19. <https://www.technologyreview.com/2023/11/14/1083366/google-deepminds-weather-ai-can-forecast-extreme-weather-quicker-and-more-accurately/> [↑](#endnote-ref-19)
20. <https://www.nature.com/articles/s41586-023-06185-3> [↑](#endnote-ref-20)
21. <https://arxiv.org/abs/2202.11214> [↑](#endnote-ref-21)