## Résultats

Table 6 présente nos modèles et les scores tirés de leurs rapports de classification.

Dans un souci de comparabilité et de focalisation, nous nous limitons ici aux 19 expériences réalisées sur le tableau unique model\_weatherAUS.csv et ne présentons pas l’intégralité des 71 expériences réalisées sur l’ensemble des 10 tableaux.

Table 6. Synthèse des rapports de classification triée par ordre décroissant selon l’accuracy (acc dans la table : les noms des métriques dans les en-têtes ont été raccourcis pour faciliter la mise en page).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Nº** | **Algorithme** | **Variation** | **acc** | **prec\_0** | **prec\_1** | **rec\_0** | **rec\_1** | **f1\_0** | **f1\_1** |
| 1 | Hard Voting optimisé | StandardScaler + SMOTE | 0,86 | 0,86 | 0,81 | 0,97 | 0,45 | 0,91 | 0,58 |
| 2 | RF | StandardScaler + SMOTE + ROC | 0,86 | 0,87 | 0,80 | 0,97 | 0,50 | 0,92 | 0,61 |
| 3 | XGBoost | StandardScaler + SMOTE + ROC | 0,85 | 0,84 | 0,88 | 0,99 | 0,35 | 0,91 | 0,50 |
| 4 | DNN | StandardScaler | 0,85 | 0,88 | 0,73 | 0,94 | 0,54 | 0,91 | 0,62 |
| 5 | Stacking optimisé | StandardScaler + SMOTE + ROC | 0,84 | 0,83 | 0,91 | 0,99 | 0,29 | 0,90 | 0,44 |
| 6 | TSF | sort + TSS + SMOTE | 0,83 | 0,84 | 0,75 | 0,96 | 0,37 | 0,90 | 0,49 |
| 7 | TSF | sort + TSS + StandardScaler + SMOTE | 0,83 | 0,87 | 0,64 | 0,91 | 0,54 | 0,89 | 0,59 |
| 8 | RF | StandardScaler + SMOTE | 0,83 | 0,89 | 0,67 | 0,88 | 0,70 | 0,88 | 0,68 |
| 9 | Soft Voting optimisé | StandardScaler + SMOTE + ROC | 0,82 | 0,82 | 0,94 | 1,00 | 0,22 | 0,90 | 0,36 |
| 10 | TSF | sort + TSS | 0,82 | 0,83 | 0,79 | 0,98 | 0,30 | 0,89 | 0,43 |
| 11 | TSF | sort + TSS | 0,82 | 0,82 | 0,79 | 0,98 | 0,30 | 0,89 | 0,43 |
| 12 | TSF | sort + TSS + StandardScaler | 0,82 | 0,83 | 0,77 | 0,97 | 0,31 | 0,89 | 0,44 |
| 13 | LogReg | StandardScaler + SMOTE | 0,82 | 0,91 | 0,63 | 0,84 | 0,77 | 0,87 | 0,70 |
| 14 | ROCKET | sort + TSS + tslearn | 0,80 | 0,80 | 0,92 | 1,00 | 0,15 | 0,89 | 0,26 |
| 15 | ROCKET | sort + TSS + MinMaxScaler + tslearn | 0,80 | 0,80 | 0,89 | 0,99 | 0,15 | 0,89 | 0,26 |
| 16 | KNN-DTW | sort + TSS + StandardScaler | 0,78 | 0,85 | 0,51 | 0,86 | 0,49 | 0,86 | 0,50 |
| 17 | KNN-DTW | sort + TSS + MinMaxScaler | 0,78 | 0,85 | 0,51 | 0,87 | 0,47 | 0,86 | 0,49 |
| 18 | KNN-DTW | sort + TSS + StandardScaler + RandomUnderSampler | 0,70 | 0,89 | 0,41 | 0,70 | 0,71 | 0,79 | 0,52 |
| 19 | ROCKET | sort + TSS + StandardScaler + RandomUnderSampler + tslearn | 0,25 | 0,99 | 0,23 | 0,03 | 1,00 | 0,05 | 0,38 |

L’application souple du seuil de 83% à l’accuracy, en admettant un écart de 1%, permet de séparer la table en deux parties :

* Partie A : nº 1 à 13
* Partie B : nº 14 à 19

La partie B, qui est exclusivement composée d’algorithmes issus de la **classification des séries temporelles,** peut être éliminée compte tenu des écarts extrêmement importants (> 80%) dans les scores de recall (pour ROCKET), ainsi que l’éloignement non négligeable (≥ 5%) des scores d’accuracy par rapport au seuil de 83% (pour KNN-DTW).

Dans la partie A, nous observons que les **méthodes d’ensemble** (Voting, XGBoost, Stacking) sont très bien classées et semblent bien fonctionner sur notre jeu de données.

Nous remarquons toutefois un écart important entre les deux implémentations du Voting, avec un meilleur équilibre des scores pour le Hard Voting (nº 1) comparé au Soft Voting (nº 9) : l’écart entre les scores de precision du premier est de 5% contre 12% pour le second, tandis que l’écart entre les scores de recall du premier est de 52% contre 78% pour le second.

Les approches basées sur les **arbres de décision** (RF et TSF) affichent également des performances très correctes.  
L’algorithme RF semble grandement bénéficier de l’optimisation du seuil (nº 2 versus nº 8) : cela fait augmenter son score de precision\_1 de 13% et réduit de 15% l’écart entre ses deux scores de precision. Ses scores de recall subissent toutefois un déséquilibre, l’écart augmentant de 18% à 47%.  
Pour les deux meilleures implémentations de l’algorithme TSF, celle sans mise à l’échelle (nº 6) a des scores de precision plus équilibrés (écart de 9%) par rapport à celle avec mise à l’échelle (nº 7) (écart de 23%). En revanche, la situation s’inverse pour les scores de recall : l’écart est de 59% pour le premier contre 37% pour le second.

Les approches basées sur les **distances** (LogReg et KNN-DTW), quant à elles, ne semblent pas particulièrement adaptées à notre jeu de données.

Après exclusion des trois implémentations de KNN-DTW (nº 16, nº 17, nº 18), il reste LogReg (nº 13), qui a obtenu des scores corrects, mais pas exceptionnels en comparaison aux autres algorithmes que nous présentons ici. Ses scores de precision présentent notamment un déséquilibre (écart de 28%).

Enfin, pour compléter cette description des résultats, nous abordons notre algorithme basé sur les **réseaux de neurones** (DNN) (nº 4).

Comme LogReg, Il affiche des bons scores dans l’ensemble, sans en proposer les meilleurs.

Nous avons tenté d’apporter de nombreuses améliorations à l’architecture du réseau (nombre de perceptrons), aux fonctions de perte (personnalisation) et aux autres mécanismes de contrôle (callbacks), mais les scores plafonnent et présentent notamment un déséquilibre au niveau du recall (écart de 40%).