

Comparaison des Méthodes de Classification

Comparaison des Performances des Modèles

Classe (effectif)	KNN	LDA	QDA	Bayésien Naïf	Arbre CART	Forêt Aléatoire	Reg Log OVA	Reg Log OVO	Reg Multinom	Réseau Neurons	SVM OVA	SVM OVO
1 (10592)	78.48	63.43	54.88	61.35	74.61	82.73	63.10	66.97	66.82	78.29	68.29	70.55
2 (14165)	80.76	65.83	55.10	61.14	76.84	87.89	75.89	75.40	75.61	86.69	78.57	78.61
3 (7151)	86.36	63.08	66.29	65.66	85.03	94.55	88.81	86.78	87.34	84.55	90.63	90.35
4 (549)	63.64	48.18	48.18	60.00	67.27	64.55	26.36	34.55	30.91	70.00	20.00	21.82
5 (1899)	77.89	47.11	50.26	46.32	69.21	71.84	16.58	25.79	24.21	83.95	33.16	34.47
6 (3473)	73.05	52.45	48.27	44.81	74.06	78.67	27.67	40.06	35.30	86.31	38.33	38.62
7 (4102)	93.79	80.88	80.88	79.29	90.86	94.28	81.36	80.39	79.66	92.33	82.10	81.49
Total (41931)	81.42	64.04	58.60	61.70	78.35	86.55	68.07	69.99	69.54	84.38	72.22	72.80

Meilleures Performances Globales

- La Forêt Aléatoire est le meilleur modèle global avec 86.55% de précision moyenne.
- Le Réseau de Neurons suit de près avec 84.38%, prouvant l'efficacité des méthodes d'apprentissage profond.
- Le KNN (81.42%) et l'Arbre CART (78.35%) sont également compétitifs.

Analyse des Classes Minoritaires (4, 5, 6)

- Les classes peu représentées sont souvent mal classées.
- Le Réseau de Neurons (70.00%, 83.95%, 86.31%) est le meilleur classifieur pour les classes à faibles effectifs.
- La Forêt Aléatoire (64.55%, 71.84%, 78.67%) suit, avec une robustesse élevée.
- L'Arbre CART (67.27%, 69.21%, 74.06%) est également robuste.

Méthodes les Plus Faibles :

- Les SVM et Les Régressions Logistiques (OVA et OVO) sous-performent sur les classes minoritaires (moins de 40% pour certaines).
- QDA, LDA et Bayésien Naïf sont globalement moins efficaces (58.60% et 61.70% respectivement).

Méthodes Paramétriques vs Non-Paramétriques

Les modèles non-paramétriques surpassent largement les modèles paramétriques : - La Forêt Aléatoire (86.55%), le Réseau de Neurons (84.38%), KNN (81.42%) et l'Arbre CART (78.35%) dominent le classement. - À l'inverse, LDA (64.04%), QDA (58.60%) et Bayésien Naïf (61.70%) sont nettement moins performants.

Les modèles non-paramétriques sont plus flexibles et capturent mieux des structures complexes dans les données, tandis que les modèles paramétriques reposent sur des hypothèses restrictives.

Multiclasse Natif vs Adapté (OVA/OVO)

Les méthodes multiclasse natives (comme Forêt Aléatoire, Arbre CART, Réseau de Neurons) ont des performances meilleures que les modèles binaires adaptés OVA et OVO.

- La Forêt Aléatoire (86.55%) et le Réseau de Neurons (84.38%), qui sont naturellement adaptés au multiclasse, surpassent les modèles SVM OVA (72.22%) et SVM OVO (72.80%), ainsi que les régressions logistiques OVA et OVO.
- Les méthodes binaires adaptées (OVA et OVO) peinent surtout sur les classes minoritaires, avec des scores très faibles (ex. SVM OVA : 20.00% sur la classe 4 !).

Explication :

- OVA force une classe unique à se démarquer contre toutes les autres.
- OVO compare les classes deux à deux, ce qui est sous-optimal pour des classes déséquilibrées.

Résumé Final

- Les modèles non-paramétriques sont les meilleurs grâce à leur flexibilité et leur adaptation aux classes déséquilibrées.
 - Les méthodes multiclassées natives (Forêt Aléatoire, Réseau de Neurones, Arbre CART, KNN) dominent les modèles binaires adaptés (SVM..).
 - Si les classes minoritaires sont importantes, privilégiez Forêt Aléatoire, Réseau de Neurones ou Arbre CART.
 - Les modèles OVA/OVO ne sont pas adaptés aux jeux de données avec des classes déséquilibrées.
-