

1. Attrition (Classification)

Objectif : Repérer les marins à risque de départ pour déclencher un soutien (moral, famille, affectation).

Modèles utilisés : Régression logistique, Arbre de décision, Random Forest, KNN.

KPI analysés : PR-AUC (car classes déséquilibrées), F1 (classe départ), seuil optimal basé sur le coût métier (faux positifs vs faux négatifs).

Résultats : La Régression Logistique optimisée est le meilleur modèle avec PR-AUC ≈ 0.62 . Elle obtient un rappel élevé (84%) mais une précision plus faible (30%). Cela veut dire qu'on détecte la majorité des départs, quitte à déclencher des alertes supplémentaires.

2. Segmentation (Clustering)

Objectif : Créer des groupes d'attention pour adapter les actions RH selon le profil (jeunes scientifiques, commerciaux expérimentés, managers seniors...).

Modèles utilisés : KMeans (K=2 et K=5), visualisation avec PCA.

KPI analysés : Silhouette (séparation entre clusters), méthode du coude (équilibre complexité/variance).

Résultats : Avec K=2 : distinction simple jeunes scientifiques vs commerciaux expérimentés. Avec K=5 : segmentation plus fine (jeunes scientifiques, commerciaux milieu de carrière, directeurs R&D; seniors, managers seniors). K=5 apporte un storytelling plus riche et utile aux RH.

3. Stabilité (Régression)

Objectif : Estimer l'ancienneté (YearsAtCompany) pour comprendre ce qui retient les marins et anticiper les parcours.

Modèles utilisés : Régression Linéaire, Arbre de décision, Random Forest.

KPI analysés : MAE (erreur moyenne en années), R^2 (qualité explicative), MAPE (erreur relative). Explicabilité via SHAP et PDP.

Résultats : Le Random Forest est le meilleur (MAE ≈ 0.85 an, $R^2 \approx 0.93$). Les variables clés sont les années avec le manager, dans le rôle, l'expérience totale et les promotions. Cela montre que la stabilité dépend surtout de la continuité managériale et des opportunités de carrière.

Conclusion générale

J'ai volontairement écarté la régression sur les salaires (MonthlyIncome) car elle n'apportait rien de pertinent pour les RH (les grilles sont déjà définies). J'ai donc concentré le projet sur 3 cas d'usage vraiment utiles : détection des départs (classification), segmentation des profils (clustering), et stabilité des carrières (régression). À chaque fois, j'ai choisi les KPI adaptés (PR-AUC, Silhouette, MAE...) et optimisé mes modèles (Grid/RandomizedSearchCV, seuils de décision, SHAP pour explicabilité). Résultat : un projet cohérent, reproductible et directement actionnable pour la Marine Nationale.