



# **Analyse des données RH de TechNova Partners**

---

Etude des déterminants possibles du  
turn-over



# Plan de la présentation

**01**

Contexte et objectifs

**02**

Schéma de travail réalisé  
et première lecture  
graphique

**03**

Feature engineering

**04**

Modélisation

**05**

Interprétation des  
résultats

**06**

Préconisations





**01**

# **Contexte et objectifs**

---



# Contexte et objectifs

Technova rencontre une problématique de turn-over importante et cherche à y remédier.

Le but est d'identifier des clés de lecture de la situation pour imaginer des leviers d'action pour limiter ce turn-over.

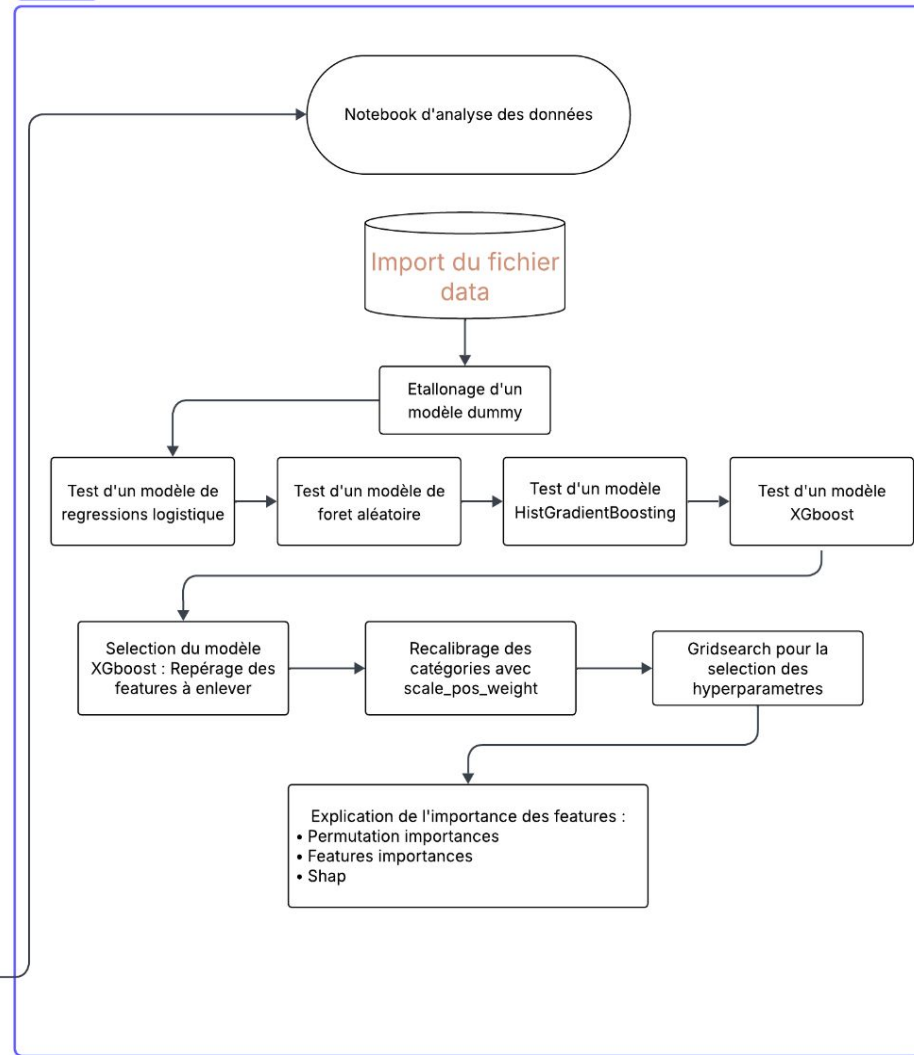
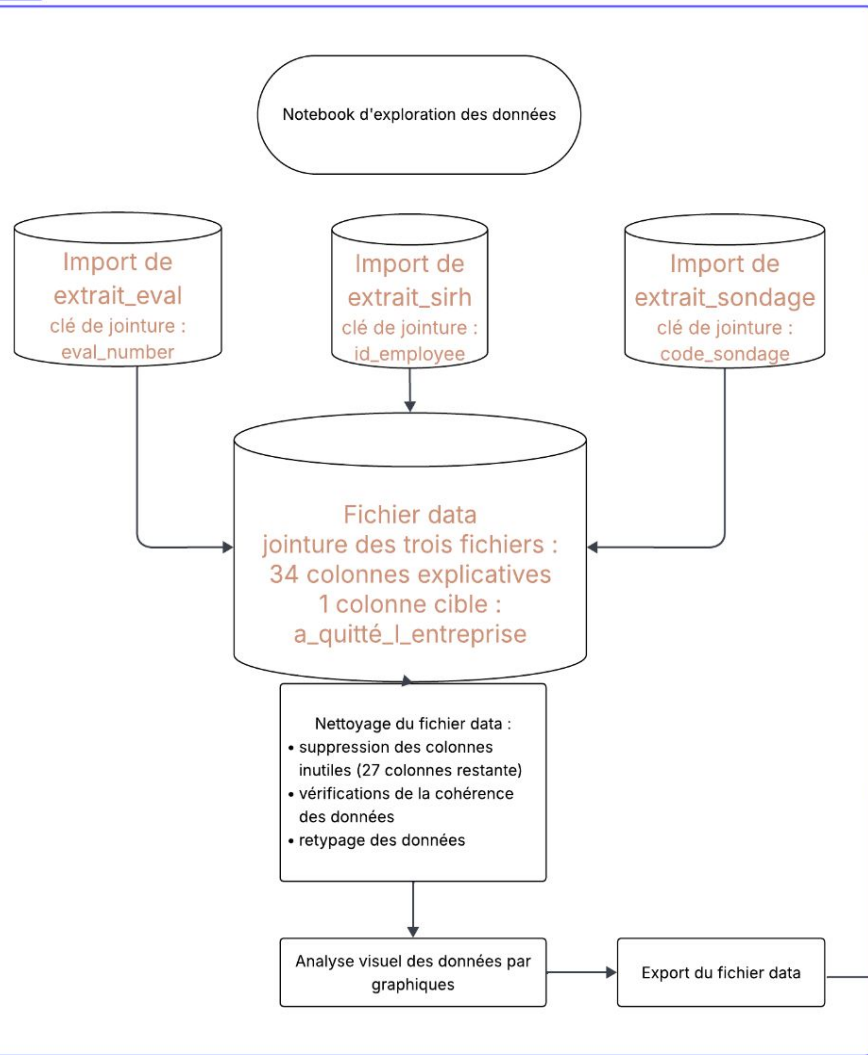
Après avoir réalisé des entretiens de sortie avec les démissionnaires, l'utilisation de données pour objectiver la situation semble souhaitable.

Nous proposons d'utiliser une approche de machine learning pour tenter d'identifier et de quantifier les facteurs associés au départ de l'entreprise.



**02**

**Schéma de travail  
réalisé et première  
lecture graphique**



# Description des données

1470 individus avec des informations sur 3 fichiers.

Une fois regroupés, ces fichiers présentent 34 colonnes explicatives possibles et une colonne cible.

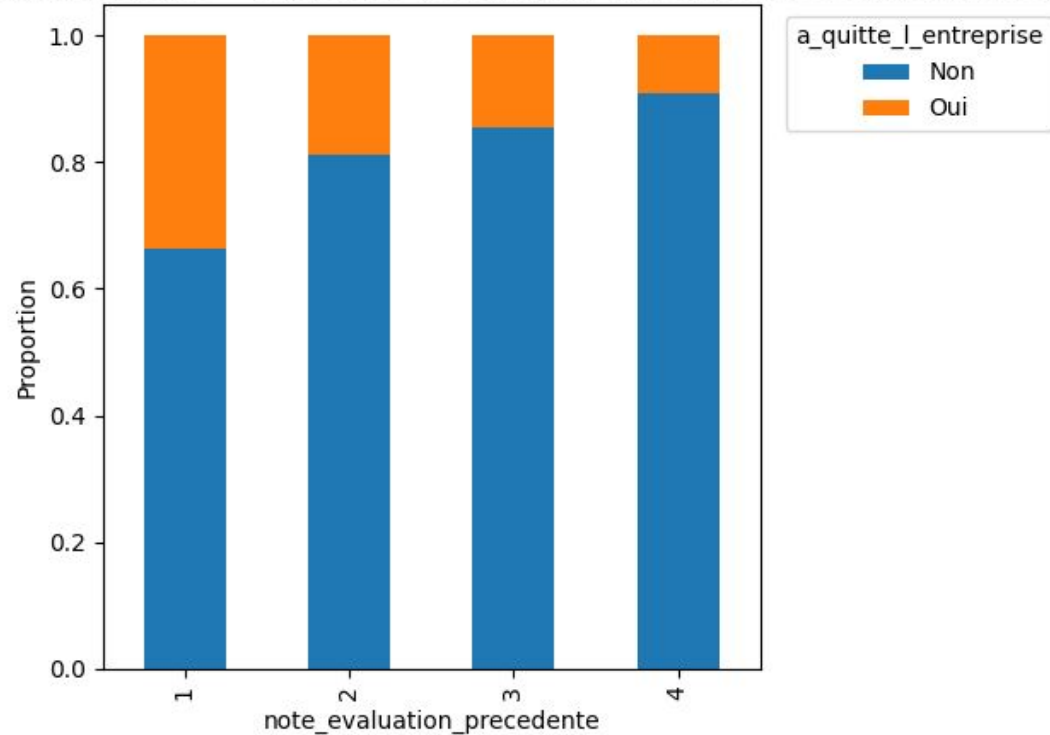
Après suppression de colonnes inutiles il reste 27 colonnes exploitables pour tenter d'expliquer le positionnement sur la colonne cible "a\_quitte\_l\_entreprise".

La variable cible présente les résultats suivants :

- Encore dans l'entreprise 1233 (  $\approx 84\%$  )
- A quitté l'entreprise 237 (  $\approx 16\%$  )

# Analyse graphique

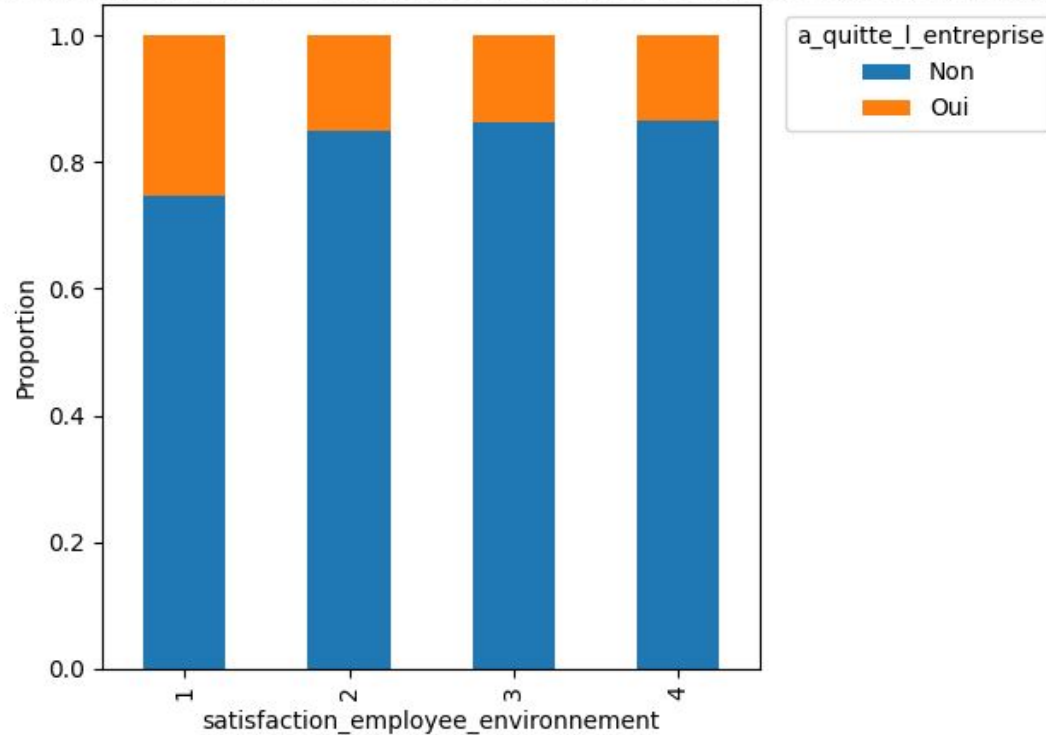
note\_evaluation\_precedente — répartition de a\_quitte\_l\_entreprise (discret, normalisée)





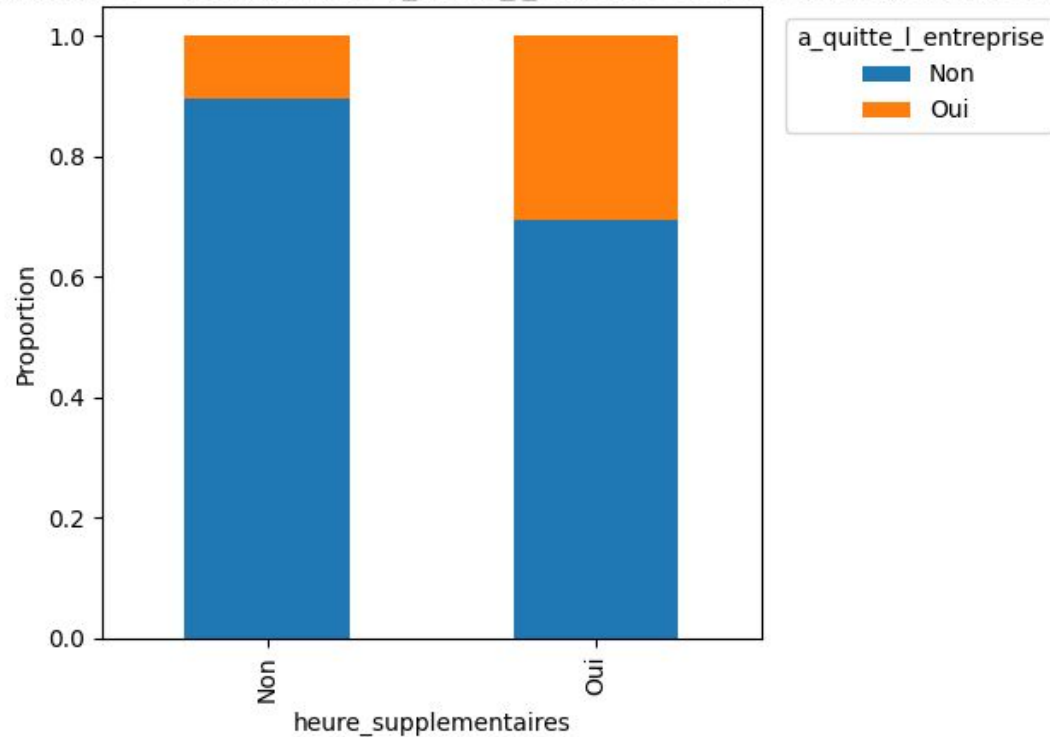
# Analyse graphique

satisfaction\_employee\_environnement — répartition de a\_quitte\_l\_entreprise (discret, normalisée)

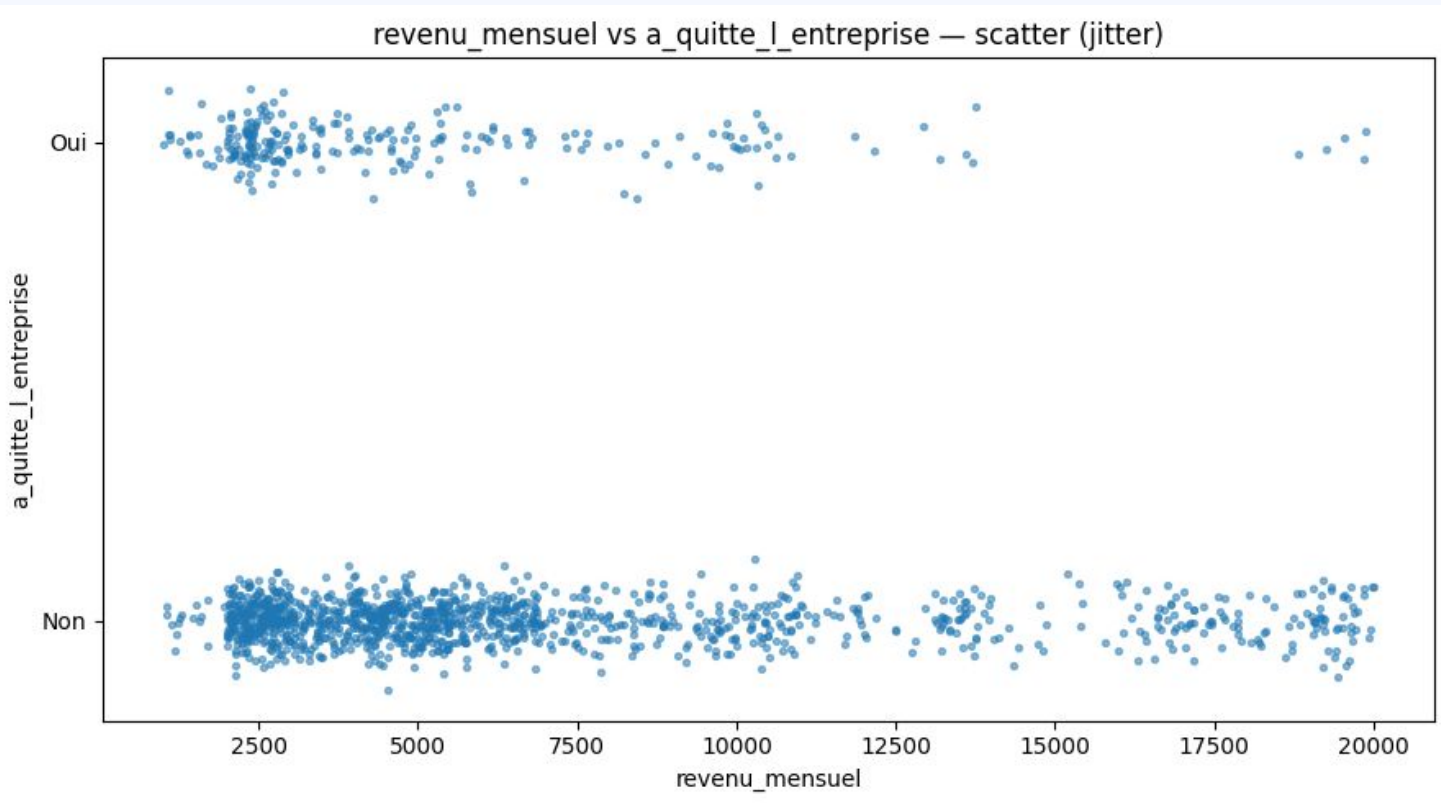


# Analyse graphique

heure\_supplementaires — répartition de a\_quitte\_l\_entreprise par modalité (normalisée)



# Analyse graphique





**03**

**Feature**

---

**engineering**



# Feature engineering

Feature	Type	Intuition métier
delta_note_evaluation	Numérique	Variation récente de l'évaluation (signal de décrochage si $> 0$ ).
perf_degrade_flag	Flag (0/1)	Dégradation de la perf (alerte simple).
perf_degrade_niv	Numérique	Intensité de la dégradation (zéro si stable/hausse).
nb_annee_hors_entreprise	Numérique	Poids de l'expérience acquise ailleurs.
Ratio_dans_et_hors_entreprise	Numérique	Mix interne/externe (profil carrières).
Hors_entreprise_majoritaire	Flag (0/1)	Expérience externe prédominante.
Ecart_nb_annee_sur_poste	Numérique	Ancienneté non passée sur le poste actuel (mobilité potentielle).

# Feature engineering

Feature	Type	Intuition métier
sat_mean	Numérique	Niveau global perçu du contexte de travail.
sat_min	Numérique	Point de friction principal.
sat_std	Numérique	Hétérogénéité des satisfactions (conflits localisés).
sat_low_flag	Flag (0/1)	Signal de contexte défavorable récurrent.
revenu_vs_dept_med	Numérique (ratio)	departement)`
revenu_vs_niveau_med	Numérique (ratio)	niveau_hierarchique_poste)`
heure_supp_flag	Flag (0/1)	Charge/effort supplémentaire.
long_commute_flag	Flag (0/1)	Contrainte de trajet (fatigue/attrition).

# Feature engineering

Feature	Type	Intuition métier
formations_par_an	Numérique	Intensité de développement (apprentissage).
recent_promo_flag	Flag (0/1)	Reconnaissance récente (ancrage).
pee_participation_flag	Flag (0/1)	Engagement financier/attachement.
pee_participation_2plus	Flag (0/1)	Engagement renforcé (récurrence).
is_manager_flag	Flag (0/1)	Responsabilités managériales (rôle structurel).
a_connu_mvmt_interne	Flag (0/1/NaN→imputé)	A déjà bougé en interne (mobilité réalisée).
tenure_ratio_current_post	Numérique	Part de la carrière passée sur le poste actuel (risque d'usure/stagnation si très élevé).



**04**

# Modélisation

---





# Modélisation

Test de 5 modèles pour la modélisation :

- 1) Modèle étalon dummy
- 2) Régression logistique
- 3) Forêt Aléatoire
- 4) HistGradientBoosting
- 5) XGboost

Au **seuil égal**, **HGB** (4) a le **meilleur F1+** de justesse, mais **XGBoost** offre le **meilleur F1 macro** et **réduit les faux positifs** (donc la charge opérationnelle).

Le **choix final** dépend du ratio coûts **FN/FP**. **XGB** est semble être le meilleur compromis.

# Modélisation

Après le calibrage du modèle XGboost, voici les résultats du modèle :

Le réglage fin a produit un XGBoost régularisé et stable (**AUC ROC 0.819, AP 0.564**).

Deux manières d'aborder les paramètres du modèle à définir selon votre réalité métier :

**Rappel-cible ( $\tau \approx 0.148$ )** : ne pas rater  $\rightarrow$  Recall 0.81, 103 alertes pour 38 vrais départs.

**F1-max ( $\tau \approx 0.223$ )** : équilibre action/impact  $\rightarrow$  F1 0.566, précision 0.485.

# Modélisation

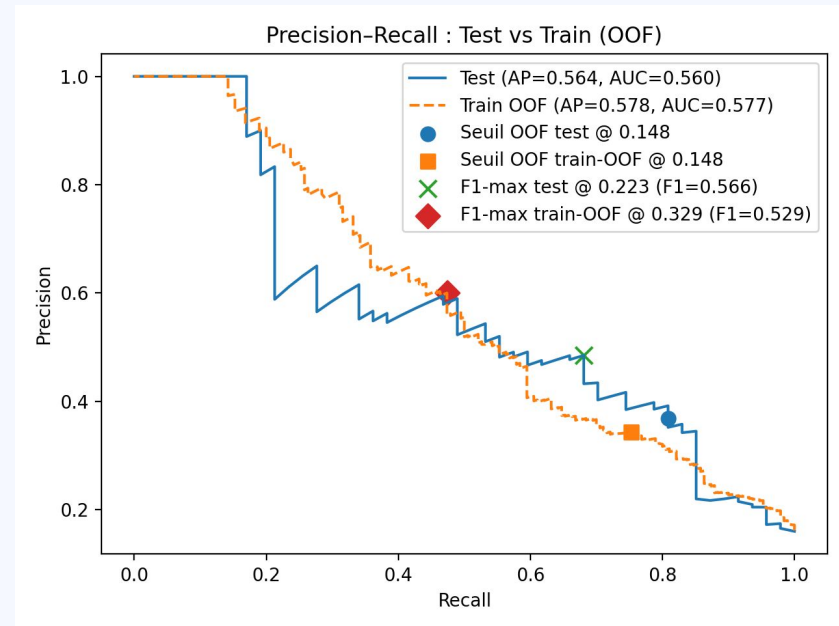
**AP (qualité du tri) – Test : 0,56 | → vrai gain.**

**Courbes Test / Train-OOF proches** → le modèle **généralise bien** (peu d'overfit).

**Seuil “rappel-cible” (0,148) : Recall  $\approx$  0,81, Precision  $\approx$  0,37** → on détecte large.

**Seuil “F1-max” (0,223) : Recall  $\approx$  0,66, Precision  $\approx$  0,48** → alertes plus ciblées.

**Conclusion** : le seuil se choisit selon le **métier** (ne pas rater de départs vs limiter le volume d'alertes).



# Résultat d'un test de détection des départs

Le modèle classe ses prédictions dans une matrice de confusion permettant d'évaluer la pertinence du modèle sur 294 personnes.

Vrai négatif (le modèle prédit reste et l'employé est réellement resté) : **182**

Faux positif (le modèle prédit part et l'employé est finalement resté) : **65**

Faux négatif (le modèle prédit reste et l'employé est finalement parti) : **9**

Vrai positif (le modèle prédit parti et l'employé est réellement parti) : **38**

Dans ce test, le modèle aurait permis d'intervenir auprès de 80,8 % des départs.

Dans cet échantillon, la taux de départs est de près de 16%. Une intervention renforcée sur les cas repérés pourrait réduire le taux des départs jusqu'à 3 % si l'ensemble des personnes identifiées était finalement restées.

Il faut aussi noter que la part importante de faux positifs peut être due à une part de personnes restées mais qui sont en fragilité et à risque pour le départ



**05**

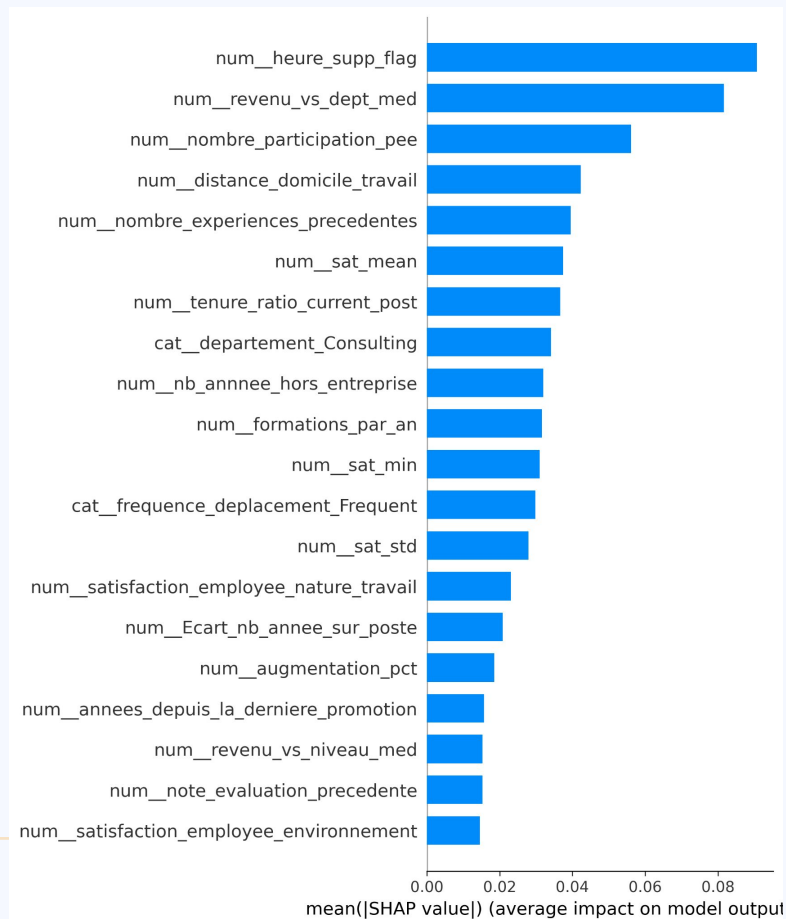
# **Interprétation des résultats**

---



# Features ayant le plus d'impact sur le modèle

Le modèle nous montre que les principaux points de vigilance à soulever portent sur les heures supplémentaires et l'équité salariale relative au sein des départements.

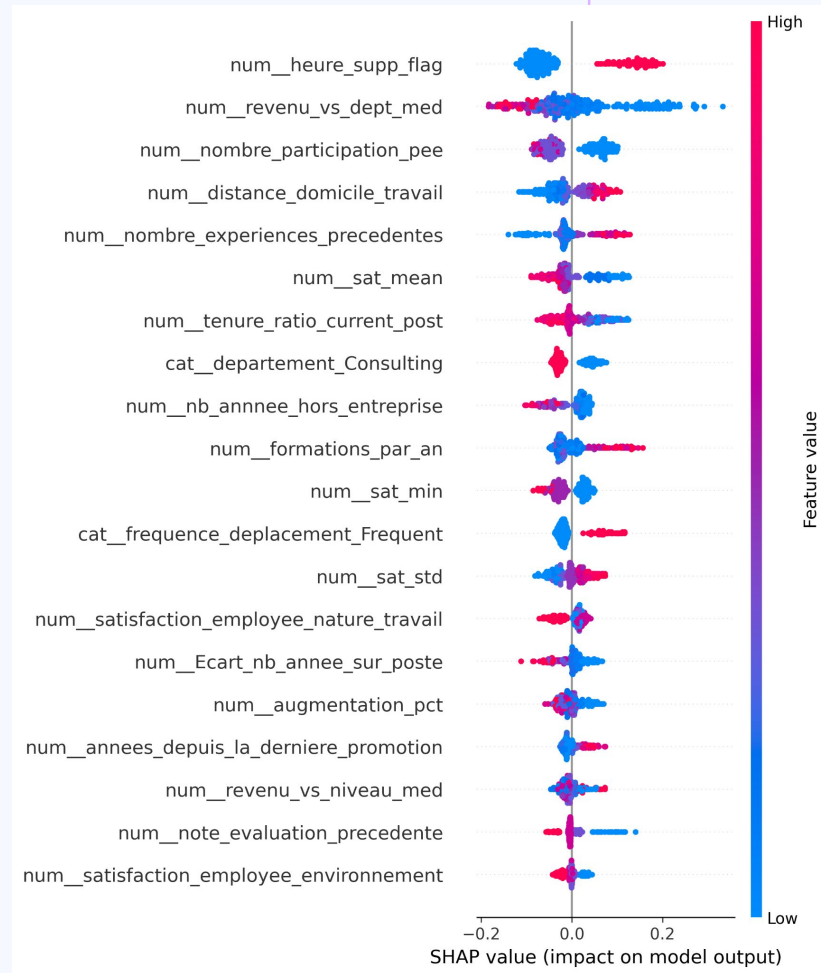


# Lecture de l'importance des features

Les principaux risques sont le suivants :

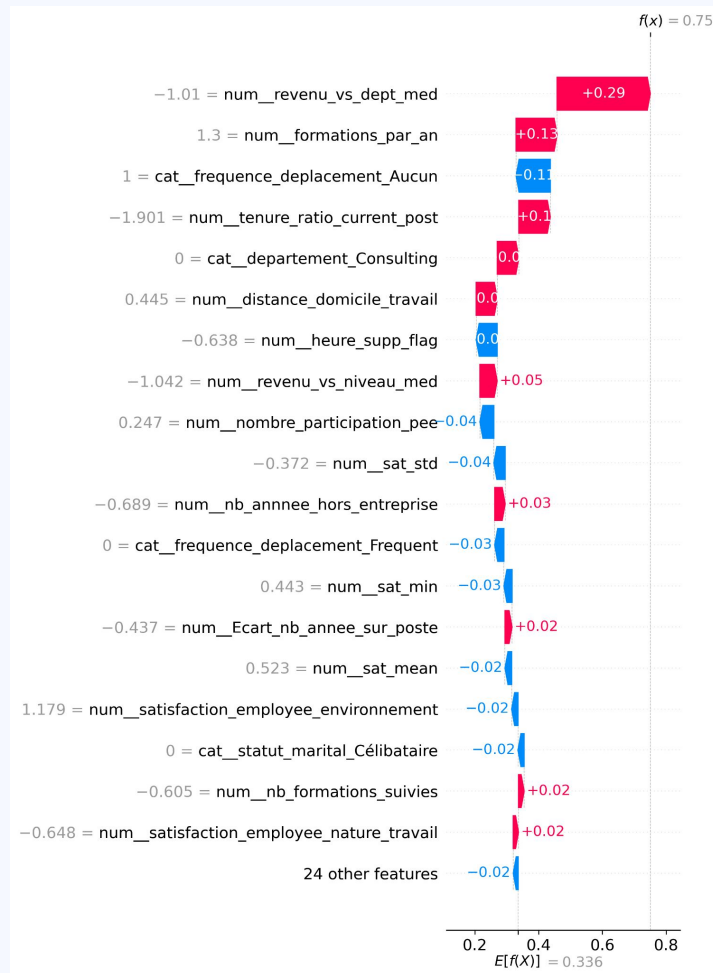
- 1) **risque de fatigue professionnelle** surtout quand il y a sur-engagement (heures sup, trajets longs),
- 2) **déséquilibre salariale relatif** (sous la médiane d'équipe),
- 3) **faible interressement financier** (PEE),
- 4) **Insatisfaction.**

Il faut aussi noter une potentielle problématique à investiguer sur le pôle consulting



# Exemple de l'impact des features pour une personne

Pour cette personne, le **salaire relatif bas** est le premier contributeur au risque, devant la **forte intensité de formation** et une **ancienneté faible** sur le poste.







**06**

# **Préconisations**

---



# Préconisations

- 1) Travailler sur la fatigue professionnelle, notamment sur les heures supplémentaires :
  - Tenter de limiter les heures supplémentaires
  - Chercher la limite sur les heures supplémentaires
  - Investiguer l'utilisation du télétravail dans la possible limitation des déplacements
- 2) Travailler sur l'équité salariale au sein des départements :
  - Investiguer la possibilité de lisser les différences salariales
  - Travailler autour de la justice organisationnelle et l'explicitation des choix salariaux
- 3) Travailler sur l'engagement des nouveaux salariés :
  - Investiguer l'impact des formations
  - Travailler sur la satisfaction générale et l'investissement
- 4) Mener une investigation particulière sur le pôle Consulting

# Préconisations

Pour suivre l'évolution des données, il serait bon de continuer le recueil des données pour vérifier si les actions mises en places ont eu un impact, mais aussi pour tenter de repérer des dynamiques qui ne sont pas encore ressortie avec le nombre d'individus dans les données.

Au vue des différents résultats, il pourrait être intéressant de récupérer des données supplémentaires pour affiner les préconisations, comme le nombre d'heures supplémentaire réalisées.

The background features a light blue gradient with abstract circuit-like patterns. Purple and orange lines, some straight and some curved, crisscross the frame. Small circles, some solid and some hollow, are placed at various points along these lines, resembling nodes or components in a circuit. The overall aesthetic is clean and modern, with a technical or digital feel.

# **Merci pour votre attention**

---

Je suis à l'écoute de vos questions