

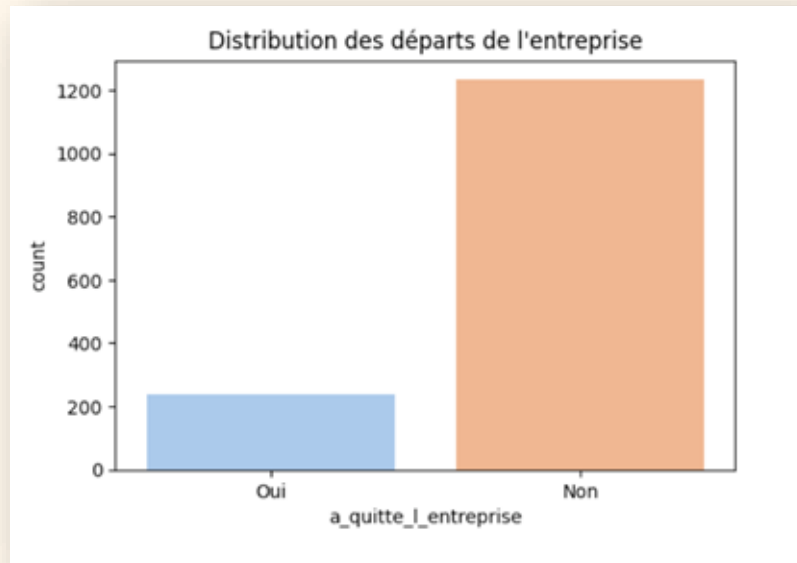
# Analyse et Prédiction de l'Attrition des Employés

Un projet de classification pour TechNova Partners

**EXIT**



# Le constat : **16 %** de nos talents ont quitté l'entreprise.



## Une Perte significative

Sur un total de 1470 employés examinés, 237 ont quitté l'entreprise. Ce taux d'attrition de 16 % constitue une perte notable.

Notre but est d'identifier les facteurs ayant contribué à cette attrition.

# Nos Sources de données



**Fichier SIRH :**  
données factuelles et  
démographiques  
(âge, genre, statut  
marital, revenu,  
poste, ...).  
1470 lignes  
11 colonnes



**Fichier d'Evaluation :**  
notes, niveau de  
satisfaction, charge de  
travail (heures  
supplémentaires, taux  
d'augmentation).  
1470 lignes  
10 colonnes



**Fichier Sondage:**  
Contient les infos  
privées ( enfants,  
trajet), le niveau  
d'études et  
la **variable cible**.  
1470 lignes  
12 colonnes

Dataset final: 1470 employés, 32 variables et aucune valeur manquante.

# Notre Démarche en 5 étapes



## **Exploration:**

Faire parler les données brutes

## **Préparation:**

Transformer et nettoyer les données pour les algorithmes.

## **Modélisation:**

Entraîner et mettre en compétition plusieurs modèles.

## **Evaluation:**

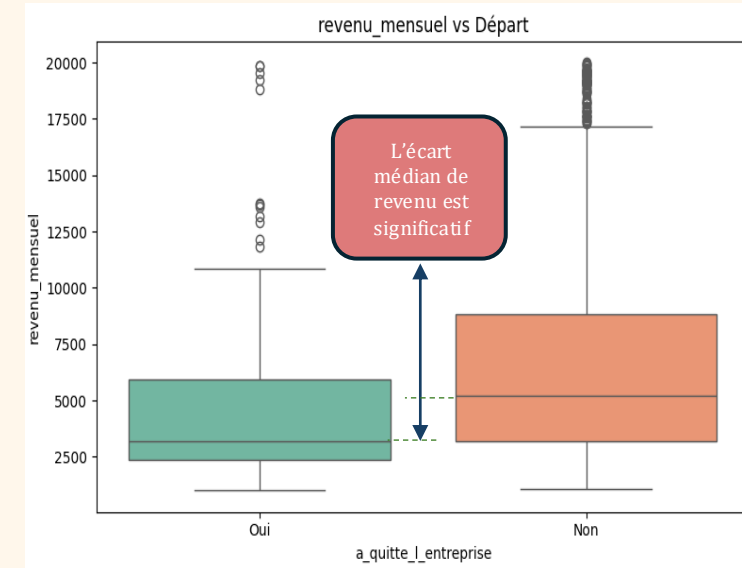
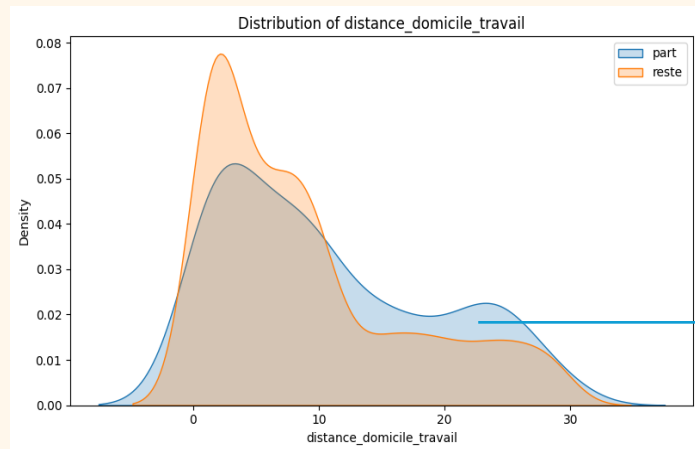
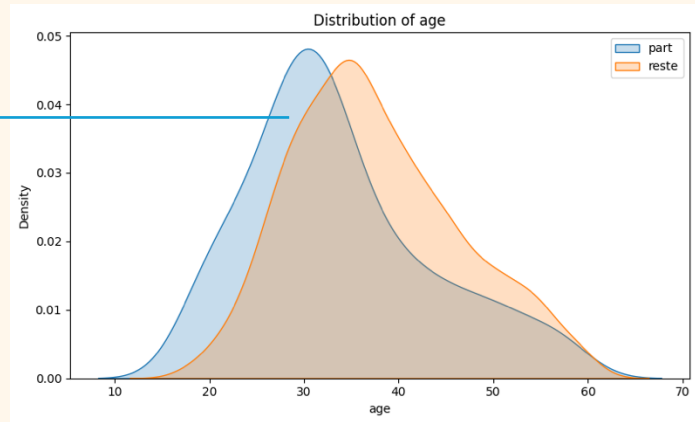
Sélectionner le modèle le plus performant.

## **Interprétation :**

Convertir les résultats en recommandations stratégiques.

# Premiers indices : Qui sont les collaborateurs qui nous quittent?

Les plus jeunes  
sont plus enclins à  
partir



Une distance domicile-travail plus élevée semble associée à un risque plus important de départ.

# Encodage et normalisation



	1	0	0,5	1,2
	0	1	3,4	-2,1
	0	0	2,8	-0,2

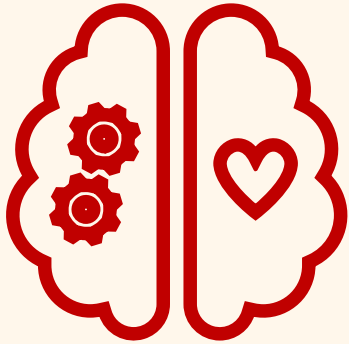
Les algorithmes ne peuvent travailler qu'avec des valeurs numériques.

Les variables textuelles ou les échelles trop différentes peuvent fausser leur apprentissage.

Pour éviter cela, nous appliquons deux étapes essentielles :

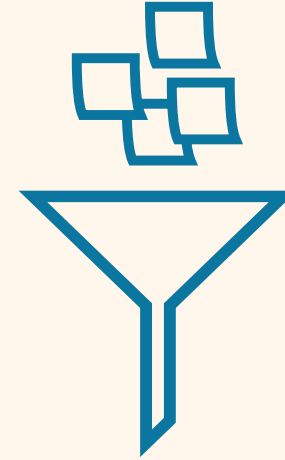
- **L'encodage**, qui consiste à transformer les variables catégorielles en valeurs numériques, notamment via le One-Hot Encoding, afin d'éviter toute notion de hiérarchie artificielle.
- **La normalisation**, qui permet de mettre toutes les variables numériques sur une même échelle pour qu'aucune ne prenne un poids excessif dans le modèle.

# Augmenter le signal et réduire le bruit



## Feature Engineering

Créer de nouvelles variables métiers pour capturer des concepts que les données brutes n'expriment pas, comme la fidélité ou la stagnation



## Elimination de la Redondance

Identifier et supprimer les variables trop corrélées qui apportent la même information et peuvent perturber l'apprentissage du modèle



# Analyse de Corrélation

Exemple : `revenue\_mensuel` et  
`niveau\_hierarchique\_poste`  
(Corrélation forte : 0,95).

Matrice de corrélation de  
Pearson

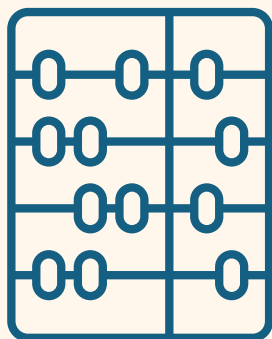
Seuil de décision >0,75

Élimination d'une des deux  
variables, en privilégiant  
celle qui apporte le  
maximum d'information.

Objectif: Supprimer les variables redondantes pour éviter la  
multicolinéarité pour faciliter le modèle



# Stratégie pour une prédiction juste



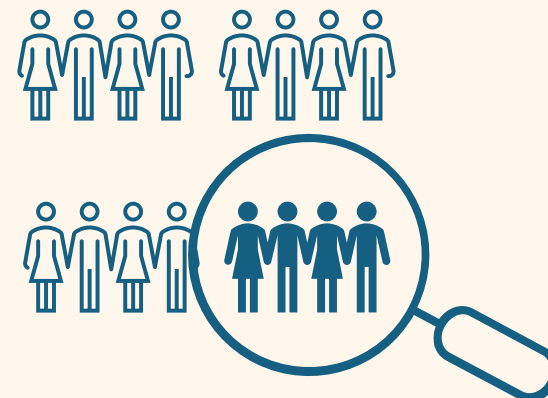
## Séparation stratifiée (stratify)

Nous assurons que la proportion de départs de restants est rigoureusement identique dans nos jeux d'entraînement et de test. Cela garantit que le modèle apprend et est évalué sur reflet fidèle de la réalité.



## Pondération des classes (class\_weight = « balanced »)

Nous instruisons nos modèles (Logistique, RF) de pénaliser plus lourdement les erreurs de classification sur la classe minoritaire (les départs). Chaque erreur sur un « départ » coûte plus cher au modèle.



## Ajustement du poids Positifs (scale\_pos\_weight)

Pour notre modèle le plus performant, XGBoost, nous utilisons une technique spécifique qui augmente l'importance totale de la classe positive (les départs) pendant l'entraînement, forçant le modèle à y prêter une attention particulière.

# La quête du modèle le plus performant

Modeles	Précision (départ)	Récall (départ)	F1-Score (global)	AUC
Régression Logistique	0.41	0.72	0.81	0.82
Random Forest	0.43	0.57	0.82	0.79
<b>XGBoost</b> 🏆	<b>0.46</b>	<b>0.66</b>	<b>0.83</b>	<b>0.79</b>



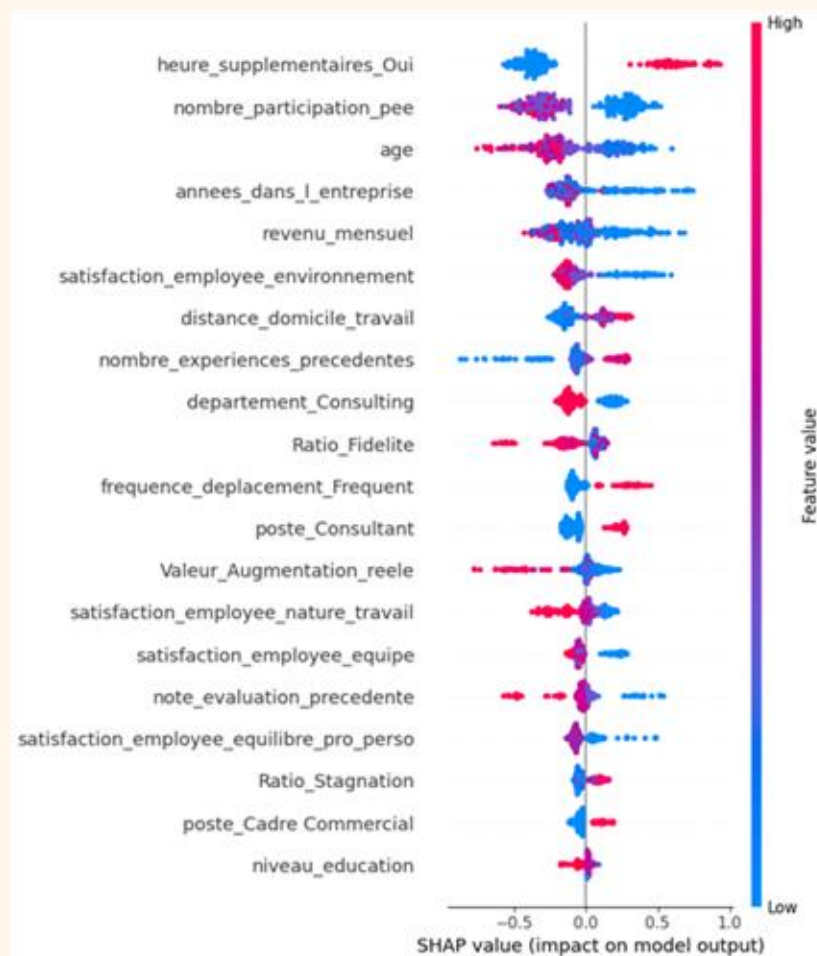
## Notre choix: Le modèle XGBoost

Il offre le meilleur équilibre entre la précision et notre capacité à détecter les départs.

Notre modèle identifie correctement 2 départs potentiels sur 3.

# Principaux facteurs clés de l'attrition

Pour comprendre la logique interne de notre modèle XGBoost, nous avons utilisé l'analyse des valeurs SHAP.



**Heure supplémentaire:** facteur N°1



**Participation au PEE**



**Âge :** Les plus jeune sont plus enclin à partir.

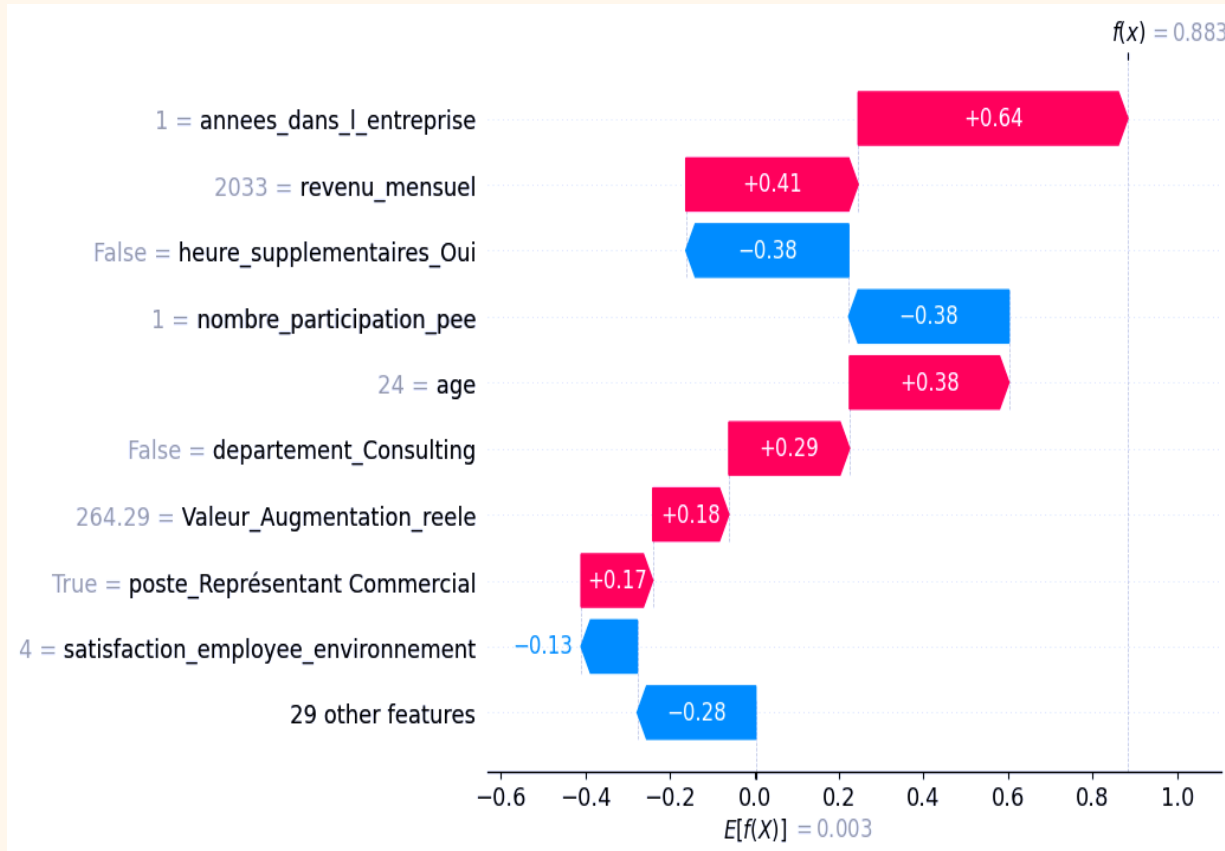


**Ancienneté:** Les collaborateurs avec peu d'ancienneté sont plus à risque.



**Revenu Mensuel:** Un faible revenu est un levier de départ.

# L'exemple d'un collaborateur à risque



Ce graphique décompose le risque de départ pour un collaborateur spécifique.

## Facteurs de Rétention:

Pour cet individu, le fait de ne pas faire d'heures supplémentaires et de participer au PEE sont des points positifs qui réduisent son risque .

## Facteurs d'attrition:

Cependant, sa faible ancienneté, son bas salaire et son jeune âge augmente très fortement son score de probable attrition à 0,88.

# Recommandations

## Pistes d'amélioration

### Optimisation avancée des modèles

- Tester plus finement les hyperparamètres
- Tester d'autres modèles

### Approche temporelle

- Intégrer une dimension temporelle pour prédire un départ sur une période
- Analyser l'évolution des indicateurs dans le temps

### Enrichissement des données

- Variables qualitatives issues de feedback de départ
- Variables exogènes

## Pistes d'actions pour TechNova

### Agir sur la charge de travail

- Mieux encadrer et compenser les heures supplémentaires, notamment pour les profils juniors, afin de limiter la fatigue et le désengagement.

### Valoriser la fidélité

- Mettre en place des parcours de carrière lisibles et des revalorisation salariales ciblées après 2 à 3 ans d'ancienneté.
- Promouvoir plus activement le plan d'épargne entreprise comme outil de fidélisation.

### Déployer le modèle sous la forme d'un tableau bord interactif pour les RH

Je vous remercie pour votre attention