

Livrable Bloc IA



BOSACKI Paul
MENNERON Laurine
PEREON Alexandre

GUEDJOU Hakim

Table des figures & tableaux	3
Introduction.....	4
Éthique	5
I. Situation Initiale	5
II. Pistes de solutions.....	5
III. Parties prenantes.....	6
IV. Points de vigilance	6
V. Solution retenue	6
VI. Respect de l'autonomie humaine, valeurs et droits.....	7
VII. Robustesse technique et sécurité.....	8
VIII. Confidentialité et gouvernance des données	9
IX. Transparence	10
X. Diversité, non-discrimination et équité.....	11
XI. Bien être environnemental et sociétal.....	12
XII. Responsabilité	14
Projet	15
I. État de l'art.....	15
II. EDA	17
III. Chargement et Compréhension des Données.....	19
Modélisation.....	29
I. Méthodes de validation des modèles	29
II. Identification des modèles à tester	32
III. Résultats	36
IV. Optimisation des hyperparamètres	37
Facteurs ayant le plus d'impact	39
Conseils à l'entreprise	41
I. Age des Employés	41
II. Distance Domicile-travail.....	41
III. Niveau hiérarchique dans l'entreprise	42
IV. Nombre d'entreprises précédentes	42
V. Moyenne d'heures travaillées par jour.....	43
Bibliographie Annotée	44

TABLE DES FIGURES & TABLEAUX

Figure 1 : Analyse Descriptive Employee Survey	22
Figure 2 : Analyse Descriptive Manager Survey	23
Figure 3 : Analyse Descriptive General Data	24
Figure 4 : Analyse Box Plot General Data	24
Figure 5 : Analyse Descriptive Work Info	25
Figure 6 : Analyse Box Plot Work Info	25
Figure 7 : Analyse PairPlot.....	26
Figure 8 : Analyse HeatMap	27
Figure 9: Principe de la validation croisée (exemple à 10 plis).....	29
Figure 10: Matrice de confusion et méthode de calcul des métriques associées	30
Figure 11: Classification avec la régression logistique	32
Figure 12: Schéma de la classification avec un SVC	33
Figure 13 : Schéma d'un arbre de décision	34
Figure 14: Principe des forêts aléatoires.....	35
Figure 15 : Courbes ROC pour nos 4 modèles.....	36
Figure 16: Comparaison des performances des modèles optimisés	38
Figure 17: Distribution des scores par métrique et par modèle	38
Figure 18: Comparaison des facteurs ayant le plus d'impact sur l'attrition selon les modèles	40
Figure 19: Attrition selon l'âge.....	41
Figure 20: Attrition selon la distance domicile-travail.....	41
Figure 21: Attrition selon le niveau hiérarchique.....	42
Figure 22: Attrition selon le nombre d'entreprises précédentes	42
Figure 23: Attrition selon le nombre moyen d'heures travaillées par jour	43

INTRODUCTION

Dans un contexte où le turnover des employés constitue un enjeu majeur pour les entreprises, la société pharmaceutique HumanForYou souhaite mieux comprendre les facteurs influençant le départ de ses employés et mettre en place des solutions pour améliorer leur rétention. Pour répondre à cette problématique, nous avons été contactés afin de réaliser une étude approfondie à l'aide des données fournies par l'entreprise sur ses employés. L'objectif de cette étude est d'identifier les principales raisons de ces départs ainsi que de proposer des pistes d'amélioration.

Ce rapport regroupe deux aspects essentiels du projet : l'éthique et la réalisation d'un algorithme pour répondre à la problématique. Dans la première partie, nous aborderons d'abord les considérations éthiques liées à l'analyse des données et à l'utilisation des modèles d'intelligence artificielle. Ensuite, nous expliquerons comment ces considérations ont été intégrées à notre projet et quelles actions nous avons mises en place pour respecter cet aspect éthique.

La seconde partie présentera notre méthodologie de travail pour concevoir notre algorithme, du prétraitement des données à l'élaboration des modèles, en passant par l'analyse des performances. Une fois l'algorithme réalisé, nous pourrions proposer des solutions à l'entreprise.

À travers cette approche, notre objectif est de fournir à HumanForYou une solution à la fois performante et respectueuse des principes éthiques, lui permettant ainsi d'optimiser la gestion de ses ressources humaines tout en garantissant un cadre de travail équitable et transparent.

ÉTHIQUE

I. SITUATION INITIALE

A) Contexte et objectif

L'entreprise HumanForYou est une entreprise pharmaceutique de 4 000 salariés basée en Inde. Elle rencontre de nombreux problèmes liés à son fort taux de rotation (15%). En effet, elle souffre de nombreux retards dans les projets et de coûts RH importants notamment à cause du temps nécessaire pour former les nouvelles recrues.

L'entreprise HumanForYou nous sollicite pour l'accompagner dans l'analyse de ses données RH afin de mieux comprendre les raisons d'un taux de rotation aussi élevé. L'objectif est donc de réduire ce taux afin de maximiser la stabilité et l'efficacité de l'entreprise.

B) Problématique éthique

Comment identifier et agir sur les facteurs influençant le départ des employés tout en respectant l'éthique et les intérêts de toutes les parties prenantes ?

II. PISTES DE SOLUTIONS

Afin de s'assurer que le projet respecte des normes éthiques, nous proposons de suivre les actions suivantes :

- Garantir la transparence dans la collecte et l'utilisation des données
- Assurer l'équité dans l'accès aux programmes de développement
- Respecter la vie privée et le droit à la déconnexion
- Maintenir un dialogue constant avec toutes les parties prenantes

Pour répondre au besoin de l'entreprises, plusieurs pistes semblent pertinentes :

- Analyse prédictive des départs à partir des données RH fournies par HumanForYou.
- Mise en place d'un programme de développement professionnel
 - Formation continue
 - Accompagnement
 - Plan de carrière personnalisé
- Amélioration des conditions de travail
 - Flexibilité du travail
 - Avantages sociaux adaptés

III. PARTIES PRENANTES

Internes

- Employés actuels
- Direction
- Service RH
- Managers intermédiaires
- Actionnaires

Externes

- Clients
- Fournisseurs
- Candidats potentiels
- Familles des employés
- Communauté locale
- Secteur pharmaceutique

IV. POINTS DE VIGILANCE

Concernant ce projet, plusieurs points de vigilance éthiques sont à relever :

- **Protection des données personnelles** : La protection des données personnelles est un enjeu majeur en 2025. En effet, avec l'essor du numérique et l'intensification des échanges de données, les risques liés aux atteintes à la vie privée, aux cyberattaques et à l'exploitation abusive des informations personnelles se multiplient. Les réglementations, telles que le RGPD en Europe, imposent des obligations strictes aux entreprises pour garantir la sécurité et la transparence dans le traitement des données. En Inde, c'est la loi DPDP 2023 (« Digital Personal Data Protection Act 2023 ») qui s'applique pour protéger les données personnelles.
- **Respect de la diversité culturelle** : En effet, l'entreprise HumanForYou est une entreprise indienne et il convient de s'adapter aux normes et/ou conventions indiennes comme le DPDP 2023 pour la protection des données.
- **Pression potentielle sur les employés** : Si l'équilibre entre surveillance et confiance n'est pas respecté, les employés pourraient ressentir une forme de pression. L'objectif de ce projet est de mieux comprendre ce qui les pousse à partir pour pouvoir améliorer leur ressenti.

V. SOLUTION RETENUE

Nous avons choisi de mettre en place l'analyse prédictive des départs pour mieux comprendre les causes des départs. Cela permettra également d'aider les employés susceptibles de partir prochainement à se sentir mieux dans l'entreprise. Ainsi, une démarche d'amélioration continue sera mise en place.

VI. RESPECT DE L'AUTONOMIE HUMAINE, VALEURS ET DROITS

A) Généralités

Le premier axe majeur de l'éthique en IA réside dans l'importance de garantir que les systèmes d'IA respectent les valeurs humaines fondamentales, notamment l'autonomie et les droits des individus. En effet, il est essentiel que les décisions automatisées de l'IA soient explicables et que les situations où les utilisateurs se retrouvent contraints par des choix biaisés ou non transparents soient évitées.

En effet, l'un des risques majeurs liés à l'IA réside dans son influence sur les choix humains. Ainsi, toute assistance apportée par l'IA peut permettre d'aider à la prise de décision mais ne doit pas contrôler la décision finale de l'utilisateur. Afin de garantir un choix libre et éclairé, le système ne doit dissimuler aucune information. De plus, l'humain doit toujours garder à l'esprit que les solutions proposées par l'IA peuvent être altérées ou inexactes.

Nous pouvons résumer ce sujet en insistant sur un point essentiel : l'importance de maintenir une éthique centrée sur l'humain. En effet, aucune IA ne doit exister dans le but de restreindre les libertés des individus. Cela est d'autant plus crucial dans des domaines comme la sécurité ou le secteur militaire. L'un des exemples les plus marquants aujourd'hui concerne la surveillance de masse en Chine, un système largement critiqué.

B) Application sur notre projet

Notre recherche portant sur l'humain et sur sa volonté de rester ou non dans une entreprise, cet axe de l'éthique est particulièrement important à considérer. En effet, si notre IA est utilisée lors du recrutement d'un nouvel employé, s'appuyer uniquement sur elle pour décider d'embaucher ou non une personne serait à la fois risqué et contraire à l'éthique. La décision humaine devra toujours primer sur le résultat de l'IA, qui ne devra servir que de conseil ou d'indication.

Pour notre projet, nous devons créer un outil d'analyse et non un outil pour faire de la prise de décision.

VII. ROBUSTESSE TECHNIQUE ET SECURITE

A) Généralités

L'IA doit être conçue pour être fiable et sécurisée. La robustesse technique implique la capacité des systèmes à fonctionner sans erreurs critiques. Il est donc essentiel de développer l'IA en intégrant des mécanismes de contrôle afin de réduire les risques de comportements inattendus ou dangereux. L'IA doit ainsi être résistante aux erreurs et aux perturbations, tout en restant fiable sur le long terme.

En termes de sécurité, plusieurs menaces doivent être anticipées :

- Résister aux attaques adversariales, c'est-à-dire aux actions visant à modifier le fonctionnement de l'IA de manière indésirable. Par exemple, un chatbot mal protégé contre ce type d'attaque pourrait finir par produire des réponses dangereuses ou biaisées.
- Réagir à l'imprévu. Bien que cela puisse sembler évident, il est crucial qu'une IA basée sur l'apprentissage soit capable de s'adapter aux situations imprévues. Toutes les possibilités doivent être envisagées afin d'éviter des conséquences potentiellement dramatiques. L'exemple le plus marquant aujourd'hui concerne les voitures autonomes : si l'IA contrôlant le véhicule n'a pas été programmée pour réagir correctement à l'apparition soudaine d'un obstacle, le risque d'accident devient considérable.

Pour faire face à ces risques de sécurité, plusieurs solutions s'offrent à nous :

- La vérification rigoureuse des systèmes, qui peut être réalisée à travers des audits, des simulations ou encore des vérifications formelles.
- L'implémentation d'un système d'alerte ou l'obligation d'une intervention humaine pour valider certaines actions de l'IA, afin d'éviter toute dérive incontrôlée.
- La sécurisation des données et des infrastructures hébergeant les informations pour prévenir tout risque de piratage. Cela peut inclure le chiffrement des données, la mise en place de systèmes redondants pour les infrastructures critiques ou encore la détection des comportements anormaux.
- L'utilisation de systèmes de filtrage, bloquant l'arrivée de données indésirables.
- L'intégration de mécanismes d'autoréparation, permettant au système de se rétablir rapidement en cas d'attaque.
- L'absence de stockage des données personnelles des utilisateurs sur des bases externes, en privilégiant leur enregistrement local sur leurs propres appareils.

B) Application sur notre projet

Notre projet n'ayant ni périmètre d'action défini ni sauvegarde des données personnelles, les dérives possibles sont donc limitées. Cependant, afin de garantir la meilleure robustesse et la meilleure sécurité, nous avons pris en compte ces aspects lors de la recherche de nos modèles afin qu'ils soient intégrés à notre choix final.

Pour renforcer encore la sécurité à notre échelle, l'ajout d'audits après la mise en production permettrait également de réduire davantage les risques.

VIII. CONFIDENTIALITE ET GOUVERNANCE DES DONNEES

A) Généralités

L'un des axes les plus importants à considérer lors de la création d'une IA basée sur des données est la sécurité et la confidentialité de ces informations. En effet, au-delà de l'aspect sécurité déjà abordé, il est également crucial de se pencher sur la protection de la vie privée. Dès que des informations privées sont utilisées, il devient essentiel de mener une enquête éthique pour garantir la confidentialité de ces données, tout en prêtant attention à l'utilisation de certaines données sensibles. Parmi ces données sensibles, on peut retrouver :

- Les données personnelles telles que le nom, l'adresse ou le numéro de téléphone, qui doivent être anonymisées.
- Les données de santé, qui sont soumises au secret médical et ne doivent donc pas être accessibles à n'importe qui.
- Les données financières, qui ne doivent logiquement pas être disponibles.
- Les données comportementales, comme les historiques de navigation, qui doivent être privées.

Toutes ces informations sont encadrées par divers documents, comme le RGPD en Europe ou l'ECPA aux États-Unis, qui imposent un encadrement strict de ces données et de leur confidentialité. Il est donc important de se concentrer sur les valeurs réellement utiles pour l'algorithme et de se débarrasser des informations sensibles non indispensables.

De la même manière que pour la sécurité, différentes actions doivent être réalisées afin de respecter les réglementations :

- Une documentation claire des algorithmes utilisant les informations, en expliquant comment et pourquoi les données collectées sont utilisées.
- L'utilisation de l'apprentissage fédéré (système de formation de modèles d'IA ne stockant pas les données sur un serveur centralisé).
- Le chiffrement et l'anonymisation des données.

B) Application sur notre projet

Pour notre projet, nous utilisons des données concernant les employés d'une entreprise, lesquelles contiennent des informations sensibles. Il est donc de notre responsabilité de veiller à ce que les personnes concernées ne soient plus identifiables.

Pour garantir l'impossibilité de réidentifier les employés, plusieurs méthodes existent, comme la confidentialité différentielle ou la modification légère de certaines colonnes.

Dans notre cas, grâce à la standardisation, il est impossible de revenir aux valeurs initiales sans connaître la moyenne et l'écart-type, rendant ainsi l'accès aux informations confidentielles compromis.

IX. TRANSPARENCE

A) Généralités

La transparence en IA signifie que les décisions prises par les algorithmes sont compréhensibles, justifiables et vérifiables par les utilisateurs et les régulateurs. Un manque de transparence dans un projet d'IA peut poser différents problèmes :

- **Manque de confiance de la part des utilisateurs.** Sans explications fournies par l'IA, ses résultats peuvent être perçus comme dangereux ou non valides.
- **Absence de responsabilité.** Sans transparence, il est difficile de désigner un responsable en cas d'erreur ou de résultat inattendu ou dangereux.
- **Difficulté d'amélioration et de supervision.** Sans une documentation claire sur l'IA, sa reprise dans le futur pourrait entraîner des incompréhensions, augmentant ainsi le risque d'erreurs.

Il y a également un important risque juridique. En effet, créer une IA non transparente va à l'encontre de réglementations telles que le RGPD (Europe) et l'ECPA (États-Unis), car cela peut mener à une utilisation incorrecte des informations. Un exemple est le risque d'utilisation de biais raciaux ou sociaux dans les IA, notamment dans le domaine de la justice. Si l'IA n'est pas transparente, identifier ces biais devient beaucoup plus difficile, ce qui pourrait rendre l'IA obsolète en raison de son manque de précision.

Pour résoudre ce problème de transparence, plusieurs solutions sont possibles :

- Comme pour la gestion de la confidentialité, une bonne documentation permet de rendre l'IA plus transparente. Il est donc important de justifier chaque action réalisée par l'IA pour la rendre compréhensible et supervisable.
- Augmenter l'explicabilité du modèle. Plusieurs options sont possibles, la plus simple étant de générer des contre-exemples à partir de notre dataset. Il s'agit de versions légèrement modifiées de celui-ci, permettant ainsi de vérifier la réaction du modèle.

B) Application sur notre projet

En termes de transparence, nous avons réalisé notre projet de telle sorte que chaque étape soit expliquée et compréhensible par une personne ne travaillant pas dessus. Cette documentation permet donc une maintenabilité accrue dans le temps, garantissant une meilleure confiance dans les résultats. Notre IA respecte ainsi les règles de transparence nécessaires afin de la rendre utilisable et digne de confiance.

X. DIVERSITE, NON-DISCRIMINATION ET EQUITE

A) Généralités

Ces trois thématiques sont des points majeurs en éthique et doivent chacune être abordées lors de la réalisation d'un projet d'IA. En effet, elles sont cruciales pour éviter qu'un modèle d'intelligence artificielle ne soit influencé par un biais lorsqu'il doit produire un résultat. Pour obtenir une IA équitable, diversifiée et non discriminatoire, il est nécessaire qu'elle intègre plusieurs concepts vitaux :

- L'équité **individuelle**. Chaque individu qui se ressemble doit être traité de manière identique, quelle que soit sa situation respective.
- L'équité entre **groupes**. Deux groupes différents doivent être traités de la même manière, quelles que soient leurs mesures statistiques globales.
- L'équité par **ignorance**. Cela correspond à l'idée de ne pas utiliser des éléments qui pourraient être perçus comme discriminatoires, afin d'éviter tout biais.
- L'égalité des **résultats**. Il est important que chaque classe démographique obtienne des résultats similaires. Cela ressemble à l'équité entre groupes, mais contrairement à cette dernière, l'égalité des résultats ne peut pas être utilisée à l'échelle individuelle. Un résultat n'est pas forcément applicable à tous, car il dépend de la situation.
- L'égalité des **chances**. Comme son nom l'indique, elle souligne l'importance de garantir que chaque personne ait la même chance d'être attribuée à une catégorie si sa situation est identique à celle d'une autre personne déjà assignée à cette catégorie.

Pour respecter ces principes, il est important de prêter attention à plusieurs détails :

- **Vérifier la qualité des données**. Il est important que la quantité de données soit représentative (en abordant toutes les possibilités), non discriminatoire (certaines anciennes informations peuvent être sexistes ou racistes) et complète.
- **Éviter tout biais**. L'algorithme ne doit pas prendre en compte de critères inéquitables, d'impacts sociaux, ni considérer les différences entre groupes de manière injuste.
- **Diversifier l'équipe développant l'algorithme** afin d'éviter qu'elle ne passe à côté de certains biais.
- **Diversifier les données d'entraînement** pour s'assurer que l'algorithme ne soit pas limité à un seul jeu de données.
- **Réguler l'IA** en imposant des standards d'équité, comme des audits.

B) Application sur notre projet

Lors de la réalisation de notre algorithme, les thèmes de la diversité, de la non-discrimination et de l'équité ont été très importants. Pour cela, nous avons veillé à ce qu'aucun favoritisme ne soit appliqué, c'est-à-dire que chaque donnée reçue soit traitée de la même manière et analysée sans biais.

De plus, nous avons également équilibré les valeurs d'attrition, toujours sans biais, afin d'obtenir de bons résultats tout en restant non discriminatoire.

L'IA étant développée pour une entreprise indienne, elle doit respecter le cadre juridique du pays. En effet, en Inde, c'est la loi DPDP 2023 (« Digital Personal Data Protection Act 2023 ») qui s'applique pour protéger les données personnelles. Cette loi étant applicable dans notre domaine, il est donc d'autant plus essentiel de rester vigilant sur ces aspects lors de la conception de notre algorithme.

XI. BIEN ETRE ENVIRONNEMENTAL ET SOCIÉTAL

A) Généralités

Les enjeux environnementaux et sociétaux sont également des points clés de l'éthique en IA. En effet, une IA mal encadrée peut engendrer divers risques, tels que l'aggravation des inégalités, une augmentation drastique de la pollution numérique ou le renforcement des fractures sociales. Cependant, l'inverse est également possible : une IA développée en tenant compte de ces enjeux peut favoriser le développement durable et améliorer la qualité de vie de la société.

Regardons en détail les enjeux environnementaux liés à l'IA. Ceux-ci peuvent être détaillés en plusieurs points :

- **L'empreinte carbone.** De nombreuses IA génératives sont en effet connues pour être de grands pollueurs, tant en termes d'énergie (GPT-3 a nécessité plusieurs centaines de mégawattheures pour son entraînement, soit autant que la consommation d'un pays comme le Danemark pendant plusieurs jours) qu'en termes d'impact carbone (certains modèles avancés d'IA consomment autant qu'un vol transatlantique).
- **L'utilisation excessive de ressources naturelles.** Pour réaliser des calculs intensifs, une quantité importante de matériel de pointe est nécessaire. Cela engendre la création d'une quantité aberrante de semi-conducteurs pour faire fonctionner certaines IA. Une autre ressource naturelle utilisée de manière excessive est l'eau. En effet, pour refroidir certaines infrastructures, des millions de litres d'eau sont nécessaires chaque jour afin d'éviter les risques de surchauffe.

Malgré des enjeux environnementaux lourds, des solutions existent :

- De plus en plus d'algorithmes sont développés pour être moins gourmands en énergie, grâce à des architectures plus efficaces (comme les modèles « pruned » ou quantifiés).
- Utiliser des énergies renouvelables pour limiter les émissions de CO2.
- Réutiliser la chaleur dégagée par les serveurs grâce à un système de recyclage thermique.
- Concevoir des IA plus petites mais mieux optimisées.
- Réutiliser le matériel informatique.

En termes d'impact sociétal, l'IA pénétrant de plus en plus dans la vie de chaque individu, il est important de prendre conscience des risques et de les éviter lors de son développement. Cependant, l'IA n'est pas uniquement négative dans ce domaine. Le problème réside dans le fait qu'une majorité des impacts positifs de l'IA comporte une part d'ombre qui doit être abordée afin d'éviter que les effets négatifs ne l'emportent sur les positifs.

Nous allons d'abord comparer les impacts sociétaux positifs et négatifs de l'IA :

- L'arrivée de l'IA dans certains secteurs, comme la logistique ou l'industrie, commence à remplacer le travail humain. À moyen terme, cela risque de créer une polarisation du marché, où les métiers nécessitant de grandes compétences seront en hausse (comme les ingénieurs en IA ou ceux travaillant dans la cybersécurité). En revanche, les emplois intermédiaires disparaîtront progressivement.
- Comme mentionné plus tôt, l'éthique reste un enjeu sociétal majeur. Bien que l'IA puisse aider à la décision dans de nombreux domaines, son aide peut rapidement se transformer en influence sur les actions ou comportements, risquant de rendre l'humain dépendant de l'IA, et non l'inverse.
- L'IA peut permettre des avancées technologiques inaccessibles autrement. Le problème réside dans l'usage que l'être humain en fait. Dans des domaines comme la médecine, l'IA peut révolutionner les diagnostics. Cependant, le développement de l'IA de reconnaissance faciale facilite la surveillance de masse ou encore la création de deepfakes.

Pour résoudre les aspects sociétaux négatifs, diverses solutions ont été envisagées et commencent à être mises en place :

- Réguler l'usage de l'IA pour éviter les dérives, comme le deepfake.
- Éduquer sur les dangers de l'IA dans les écoles et les universités.
- Encourager la collaboration entre les entreprises, les chercheurs et les gouvernements afin de créer des IA responsables.

B) Application sur notre projet

Dans notre projet, des actions environnementales et sociétales ont été réalisées. Tout d'abord, notre IA n'est pas une grande consommatrice d'énergie, réduisant ainsi l'empreinte carbone qu'elle peut produire. De plus, afin de continuer à réduire cette valeur, nous avons mis en place un barème de temps d'exécution. En effet, nous avons calculé la vitesse d'exécution de chacun des modèles analysés et, selon la réponse, nous décidons s'il est utilisable ou non, car suffisamment peu consommateur pour être exploitable sur le long terme sans risque environnemental.

En termes sociétaux, notre IA n'est pas décisionnaire mais seulement indicative, elle n'a donc pas un grand impact sociétal. Il est simplement important de vérifier qu'aucune décision ou aucun choix d'embauche ne soit réalisé à l'aide de l'IA lors de son utilisation.

XII. RESPONSABILITE

A) Généralités

Lors de la réalisation d'une IA, les personnes ou le groupe en charge du développement de celle-ci doivent en être responsables. Cela implique des responsabilités légales et politiques concernant son utilisation et les résultats qu'elle peut engendrer. Ces responsabilités sont importantes pour plusieurs raisons :

- Avoir un responsable en cas d'erreur ou de préjudice.
- Éviter que l'IA ne serve d'excuse à des actions injustes ou discriminatoires.
- S'assurer que l'IA respecte les réglementations en vigueur.

Cependant, ce système de responsabilité en est encore à ses débuts, et de nombreuses IA évoluent dans un flou juridique. Cela peut donc engendrer divers problèmes :

- Le flou en termes de responsabilité. Si un responsable n'est pas défini avant l'utilisation de l'IA, qui est responsable si une erreur subvient ? L'utilisateur ? Le développeur ?
- Les problèmes juridiques en cas de dommages.
- Les risques d'abus pouvant amener l'IA à prendre des décisions arbitraires et possiblement discriminatoires.
- L'utilisation de l'IA de manière privée et illégale pour en tirer un profit.

Pour résoudre ces problèmes, diverses solutions existent :

- La solution la plus logique reste la déclaration claire des responsables de l'IA, comme mentionné précédemment.
- Un système de supervision peut également permettre de réguler l'utilisation de l'IA et vérifier son bon fonctionnement.
- Rendre les IA plus explicables et auditées.
- Permettre un droit de recours afin de contester les actions réalisées à l'aide d'une IA.
- Établir des règles claires pour éviter tout débordement de la part des entreprises (comme la limitation de l'usage de l'IA afin d'éviter tout abus).

B) Application sur notre projet

En termes de responsabilité, nous avons décidé que l'utilisateur serait entièrement responsable des actions qu'il mettra en place grâce aux informations données par notre système. L'IA est également développé pour l'usage privé de l'entreprise, cela permet d'éviter au maximum les dérives potentielles d'une IA publique. Cependant, un ajout de supervision serait tout de même un intéressant ajout pour assurer le fonctionnement de l'IA.

PROJET

I. ÉTAT DE L'ART

Le taux de rotation des employés, ou turnover, est une préoccupation majeure pour les entreprises, car il engendre des coûts élevés liés au recrutement, à la formation et à la perte de compétences. L'intelligence artificielle (IA) est de plus en plus explorée comme un levier pour anticiper et réduire ce turnover en optimisant la gestion des ressources humaines (GRH). Cette étude analyse les systèmes d'IA déjà mis en place ainsi que leurs impacts sur la rétention des employés.

A) L'IA au service de la prédiction du turnover

L'un des principaux apports de l'IA en gestion des talents est sa capacité à analyser de grands volumes de données pour prédire les départs potentiels. Plusieurs modèles d'apprentissage automatique sont utilisés à cet effet :

- **Réseaux de neurones et forêts aléatoires** : Selon Nosratabadi et al. (2022), ces modèles sont capables d'identifier des motifs dans les comportements des employés menant à un départ (évolution des performances, absentéisme, interactions avec les managers, etc.) [7].
- **Traitement du langage naturel (NLP)** : Des systèmes d'analyse sémantique permettent d'extraire des signaux d'insatisfaction à partir des évaluations internes, des emails ou des interactions sur les plateformes collaboratives (Yircof, 2019) [11].
- **Algorithmes de régression et clustering** : Ils segmentent les employés en groupes homogènes pour identifier les profils les plus à risque (âge, ancienneté, fonction, etc.) et adapter les stratégies de rétention.

B) Amélioration de l'engagement grâce à l'IA

Au-delà de la prédiction, l'IA est utilisée pour améliorer l'engagement des employés par différents moyens :

- **Personnalisation des parcours de carrière** : L'IA peut recommander des formations et des opportunités de mobilité interne en fonction des compétences et préférences des employés (SSRN, 2024) [6].
- **Chatbots RH** : Intégrés aux plateformes d'entreprise, ces assistants répondent aux questions des employés sur les avantages, la formation ou les opportunités d'évolution, réduisant ainsi le sentiment de frustration et l'attrition.
- **Feedback en temps réel** : L'analyse des données issues d'enquêtes de satisfaction permet d'ajuster rapidement les politiques de bien-être au travail (Nosratabadi, 2022) [7].

C) Systèmes déjà mis en place

Plusieurs entreprises ont déjà adopté des systèmes basés sur l'IA pour réduire leur turnover :

- **IBM Watson Talent Insights** : Utilisé pour identifier les tendances de départ et proposer des stratégies de rétention.
- **SAP SuccessFactors** : Intègre des algorithmes prédictifs pour analyser la satisfaction des employés.
- **Workday Adaptive Insights** : Exploite l'IA pour recommander des plans de succession et des parcours de carrière personnalisés.

Ces outils ont permis de réduire le taux de rotation dans plusieurs entreprises en identifiant précocement les risques et en adaptant les politiques RH en conséquence.

D) Limites et défis

Malgré ces avancées, plusieurs défis persistent (Aschbacher, K) [2] :

- **Biais des algorithmes** : Les modèles peuvent reproduire des discriminations existantes si les données d'entraînement sont biaisées.
- **Protection des données** : L'analyse prédictive repose sur des données sensibles, soulevant des questions éthiques et légales.
- **Acceptation par les employés** : La transparence sur l'utilisation de l'IA dans les décisions RH est cruciale pour éviter une perception de surveillance intrusive.

E) Conclusion

L'IA constitue un outil prometteur pour réduire le taux de rotation des employés en permettant une meilleure anticipation des départs et une amélioration de l'engagement. Toutefois, pour maximiser son impact, les entreprises doivent veiller à une mise en œuvre éthique et à une gouvernance adaptée des données.

II. EDA

A) Objectifs du projet d'IA

Le projet "IA for HumanForYou" vise à aider l'entreprise pharmaceutique HumanForYou, basée en Inde, à réduire son taux élevé de rotation des employés, actuellement à 15 % par an. Cette situation a des impacts négatifs importants, notamment :

- **Retards dans les projets**, ce qui nuit à la réputation de l'entreprise auprès de ses partenaires et clients.
- **Augmentation des coûts RH**, en raison de la recherche et de la formation de nouveaux employés.
- **Baisse de la productivité**, les nouveaux employés nécessitant un temps d'adaptation et de formation avant de devenir pleinement opérationnels.

L'objectif principal est donc de **déterminer les facteurs influençant le départ des employés** à l'aide de techniques d'analyse de données et de modélisation IA. Ce projet permettra également de proposer des pistes concrètes pour améliorer la rétention des employés, tout en optimisant les coûts et les efforts de gestion des ressources humaines.

B) Rôle de l'EDA dans le projet

L'Exploratory Data Analysis (EDA) joue un rôle central dans ce projet d'IA pour les raisons suivantes :

- **Compréhension des données** : Identifier les caractéristiques principales des employés, leurs habitudes de travail et les variables liées à leur départ.
- **Identification des problèmes de qualité des données** : Détecter les valeurs manquantes, les incohérences ou les doublons pouvant fausser les analyses.
- **Découverte des relations entre variables** : Mettre en lumière les corrélations entre des facteurs comme le salaire, l'ancienneté, la satisfaction au travail et la probabilité de départ.
- **Visualisation des patterns** : Repérer les tendances ou anomalies dans les données grâce à des visualisations adaptées (boxplots, heatmaps, etc.).
- **Préparation pour la modélisation** : Sélectionner et transformer les variables pertinentes en vue d'une meilleure performance des modèles prédictifs.

En somme, l'EDA fournit les bases pour des décisions éclairées, garantissant que les étapes ultérieures du projet reposent sur une analyse approfondie et fiable.

C) Description générale des données

Les données fournies par HumanForYou proviennent principalement de trois sources : le service des ressources humaines, des enquêtes internes et des enregistrements de temps de travail. Elles ont été anonymisées et sont organisées en plusieurs fichiers CSV :

Données RH (*general_data.csv*)

- Informations démographiques et professionnelles, telles que l'âge, le niveau d'étude, le poste, le salaire mensuel et l'ancienneté.
- Variables clés pour l'étude, comme l'attribut *Attrition*, qui indique si un employé a quitté l'entreprise en 2016.

Évaluations des managers (*manager_survey_data.csv*)

- Scores attribués par les managers concernant l'implication des employés (*JobInvolvement*) et leurs performances annuelles (*PerformanceRating*).

Enquête sur la qualité de vie au travail (*employee_survey_data.csv*)

- Retours des employés sur trois aspects : satisfaction au travail (*JobSatisfaction*), environnement de travail (*EnvironmentSatisfaction*) et équilibre vie professionnelle/vie privée (*WorkLifeBalance*).

Horaires de travail (*in_time.zip*, *out_time.zip*)

- Enregistrements des heures d'entrée et de sortie des employés sur une période représentative de l'année 2015.

Chaque source de données contient des informations cruciales pour explorer les facteurs influençant la rétention des employés et fournir des recommandations adaptées à la direction de l'entreprise.

III. CHARGEMENT ET COMPREHENSION DES DONNEES

A) Sources et formats des données

Les données utilisées dans ce projet proviennent de plusieurs sources et sont fournies sous forme de fichiers CSV. Chaque fichier contient des informations spécifiques concernant les employés, leur historique professionnel et leur environnement de travail.

Les fichiers de données ont été extraits et rassemblés pour une analyse approfondie.

B) Chargement des données

Pour analyser ces données, elles ont été chargées à l'aide d'outils de traitement adaptés. Une première étape a consisté à les importer afin de pouvoir les explorer et en comprendre la structure. Une attention particulière a été portée à la gestion des formats, des types de données et à la présence éventuelle de valeurs manquantes.

C) Outils

Pour l'analyse des données, plusieurs bibliothèques et outils ont été utilisés :

- **pandas** : pour le chargement, la manipulation et l'analyse des données tabulaires.
- **seaborn** : pour la visualisation des données et la création de graphiques statistiques avancés.
- **matplotlib** : pour la génération de graphiques personnalisés.
- **ydata-profiling** : pour l'exploration automatique des données et la génération de rapports détaillés.

Ces outils permettent d'explorer efficacement les données, de détecter les anomalies et de visualiser les tendances pour faciliter l'interprétation des résultats.

D) Aperçu des données

Statistiques générales

- **Nombre total d'observations** : 4410
- **Nombre total de variables** :
 - Employee Survey Data : 4 variables
 - Manager Survey Data : 3 variables
 - General Data : 24 variables
 - In Time : 262 variables
 - Out Time : 262 variables
- **Total des valeurs manquantes** :
 - Employee Survey Data : 83 valeurs manquantes (0.5%)
 - Manager Survey Data : 0 valeur manquante (0.0%)
 - General Data : 28 valeurs manquantes (< 0.1%)
 - In Time : 109080 valeurs manquantes (9.4%)
 - Out Time : 109080 valeurs manquantes (9.4%)
- **Lignes dupliquées** : Aucune

Types de variables

- **Numériques :**
 - EmployeeID (identifiant unique des employés)
 - Age (âge des employés, 18-60 ans)
 - DistanceFromHome (distance en km, min: 1, max: 29)
 - MonthlyIncome (revenu mensuel, min: 10 090, max: 199 990)
 - TotalWorkingYears (expérience totale, min: 0, max: 40)
 - YearsAtCompany (années passées dans l'entreprise, min: 0, max: 40)
- **Catégorielles :**
 - JobSatisfaction (satisfaction professionnelle, 1-4)
 - WorkLifeBalance (équilibre travail-vie personnelle, 1-4)
 - JobRole (rôle professionnel, 9 catégories)
 - MaritalStatus (état civil : Marié, Célibataire, Divorcé)
 - EducationField (domaine d'étude, 6 catégories)
- **Booléennes :**
 - Attrition (le salarié a quitté l'entreprise : Yes/No)
- **Constantes :**
 - Over18 (tous les employés sont majeurs : Y)
- **Données temporelles :**
 - in_time et out_time : fichiers contenant les heures d'entrée et de sortie des employés.
 - Chaque fichier a une structure où les lignes représentent les employés via leur ID, et les colonnes correspondent aux dates spécifiques.
 - Les valeurs sont des horodatages indiquant l'heure d'entrée (in_time) et de sortie (out_time) pour chaque employé.
 - Certaines valeurs peuvent être absentes (NaN), ce qui signifie qu'aucune donnée n'a été enregistrée pour cette date spécifique.

IV. Nettoyage des Données

A) Gestion des valeurs manquantes

Le traitement des valeurs manquantes est essentiel pour garantir la qualité des analyses. Plusieurs stratégies ont été utilisées :

- **Suppression des colonnes constantes** : Les colonnes constantes ont été supprimées, car elles n'apportent aucune variation dans les données et ne contribuent donc pas à l'entraînement du modèle. Ainsi nous avons supprimé la colonne `Over18` (détermine si l'employé est majeur : toujours vrai) et la colonne `EmployeeCount` (détermine si l'employé était présent en 2015 : toujours à 1).
- **Remplissage des valeurs manquantes pour les variables catégorielles** : La valeur la plus fréquente a été utilisée pour combler les données manquantes.
- **Remplissage des valeurs manquantes pour les variables numériques** : La médiane de chaque colonne a été utilisée pour imputer les valeurs absentes.
- **Remplissage des années d'expérience** : Lorsqu'une valeur était manquante pour `TotalWorkingYears`, elle a été remplacée par la valeur de `YearsAtCompany`.

B) Détection et traitement des doublons

Aucune ligne en double n'a été trouvée dans les différentes données. Néanmoins, une vérification a été réalisée pour s'assurer qu'aucun employé n'apparaît plusieurs fois avec des informations redondantes.

C) Standardisation et conversion des types de données

- **Conversion des nombres en entiers** : Les colonnes numériques ont été converties en `int` lorsque cela était possible, en particulier si elles ne contenaient que des valeurs entières.
- **Transformation des valeurs booléennes** : La variable `Attrition` a été convertie en booléen (`True` pour "Yes" et `False` pour "No").
- **Fusion des données horaires** : Les fichiers `in_time` et `out_time` ont été restructurés pour créer une table unifiée indiquant les heures d'arrivée et de départ des employés par jour.

D) Analyse Descriptive

L'analyse descriptive des données permet d'obtenir un premier aperçu des caractéristiques principales des différentes données utilisées. Nous avons généré des statistiques de base, telles que la moyenne, la médiane, l'écart-type, ainsi que les valeurs minimales et maximales pour chaque variable.

I. Données d'enquête des employés (Employee Survey Data)

Les variables **EnvironmentSatisfaction**, **JobSatisfaction** et **WorkLifeBalance** ont été analysées pour 4410 employés.

Les données montrent que la majorité des employés sont satisfaits de leur environnement de travail, avec 1375 indiquant "Good" et 1334 "Excellent", tandis que 856 et 845 employés trouvent leur environnement "Fair" ou "Poor".

Concernant la satisfaction professionnelle, 1367 employés sont "Very Satisfied" et 1343 "Satisfied", mais 860 et 840 sont "Dissatisfied" ou "Neutral".

Enfin, pour l'équilibre travail-vie personnelle, 2698 employés le jugent "Good", 1019 "Average", et 454 "Excellent", tandis que 239 considèrent leur équilibre comme "Poor".

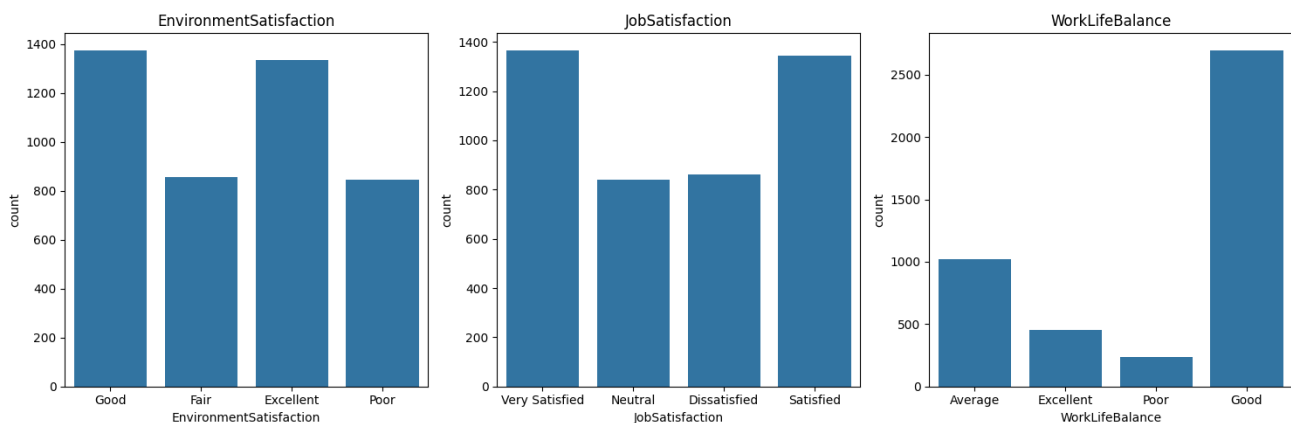


Figure 1 : Analyse Descriptive Employee Survey

2. Données d'enquête des managers (Manager Survey Data)

Pour JobInvolvement (Engagement au travail), la majorité des employés sont "Highly Engaged" (2604), suivis de 1125 employés "Moderately Engaged". Moins nombreux sont ceux qui sont "Fully Committed" (432) ou "Not Engaged" (249).

Concernant PerformanceRating (Évaluation de la performance), la majorité des employés ont une performance "Good" (3732), tandis que 678 sont évalués comme "Outstanding".

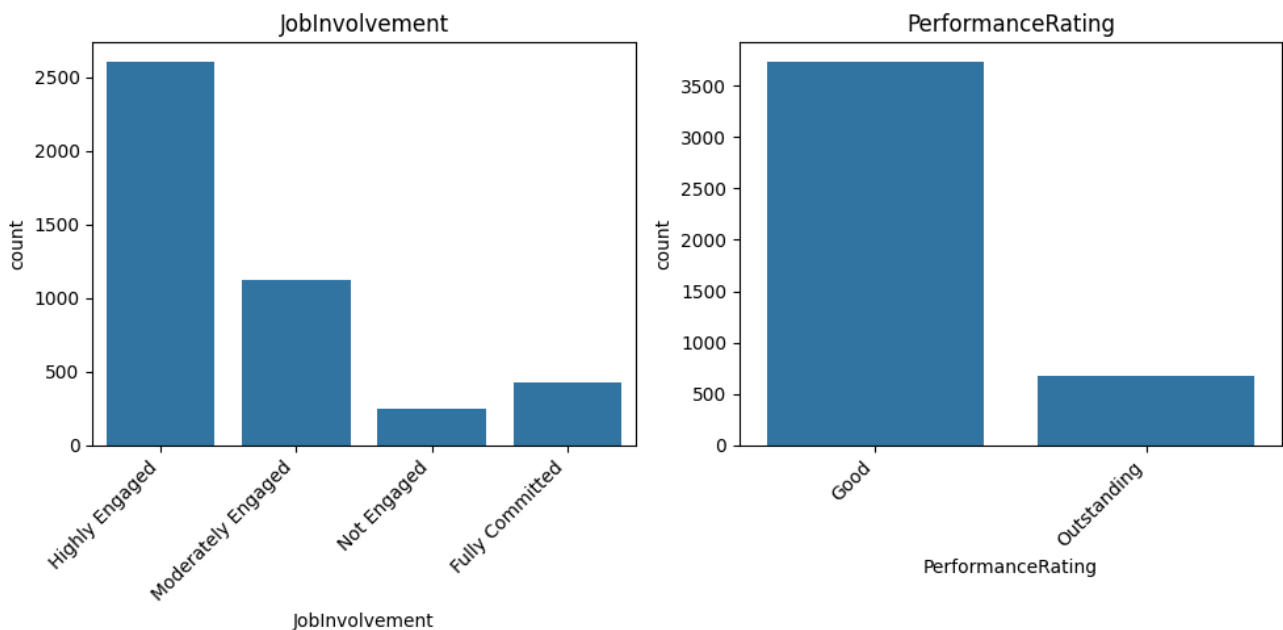


Figure 2 : Analyse Descriptive Manager Survey

3. Données générales (General Data)

L'analyse des statistiques générales met en évidence plusieurs tendances importantes :

- **L'âge moyen** des employés est de **36,92 ans**, avec un minimum de **18 ans** et un maximum de **60 ans**.
- **La distance moyenne entre le domicile et le travail** est de **9,19 km**, bien que la dispersion soit élevée avec un écart-type de **8,10 km**.
- **Le salaire mensuel moyen** est de **65 029**, avec une forte dispersion (écart-type de **47 068**), ce qui reflète une grande disparité salariale.
- **Le nombre moyen d'années travaillées** est de **11,27 ans**, avec une médiane de **10 ans**, et certaines personnes ayant travaillé jusqu'à **40 ans**.

	Age	DistanceFromHome	Education	JobLevel	MonthlyIncome	NumCompaniesWorked
count	4410.000000	4410.000000	4410.000000	4410.000000	4410.000000	4391.000000
mean	36.923810	9.192517	2.912925	2.063946	65029.312925	2.694830
std	9.133301	8.105026	1.023933	1.106689	47068.888559	2.498887
min	18.000000	1.000000	1.000000	1.000000	10090.000000	0.000000
25%	30.000000	2.000000	2.000000	1.000000	29110.000000	1.000000
50%	36.000000	7.000000	3.000000	2.000000	49190.000000	2.000000
75%	43.000000	14.000000	4.000000	3.000000	83800.000000	4.000000
max	60.000000	29.000000	5.000000	5.000000	199990.000000	9.000000

	PercentSalaryHike	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	YearsAtCompany	YearsSinceLastPromotion	YearsWithCurrManager
count	4410.000000	4410.0	4410.000000	4410.000000	4410.000000	4410.000000	4410.000000	4410.000000
mean	15.209524	8.0	0.793878	11.275737	2.799320	7.008163	2.187755	4.123129
std	3.659108	0.0	0.851883	7.780539	1.288978	6.125135	3.221699	3.567327
min	11.000000	8.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	12.000000	8.0	0.000000	6.000000	2.000000	3.000000	0.000000	2.000000
50%	14.000000	8.0	1.000000	10.000000	3.000000	5.000000	1.000000	3.000000
75%	18.000000	8.0	1.000000	15.000000	3.000000	9.000000	3.000000	7.000000
max	25.000000	8.0	3.000000	40.000000	6.000000	40.000000	15.000000	17.000000

Figure 3 : Analyse Descriptive General Data

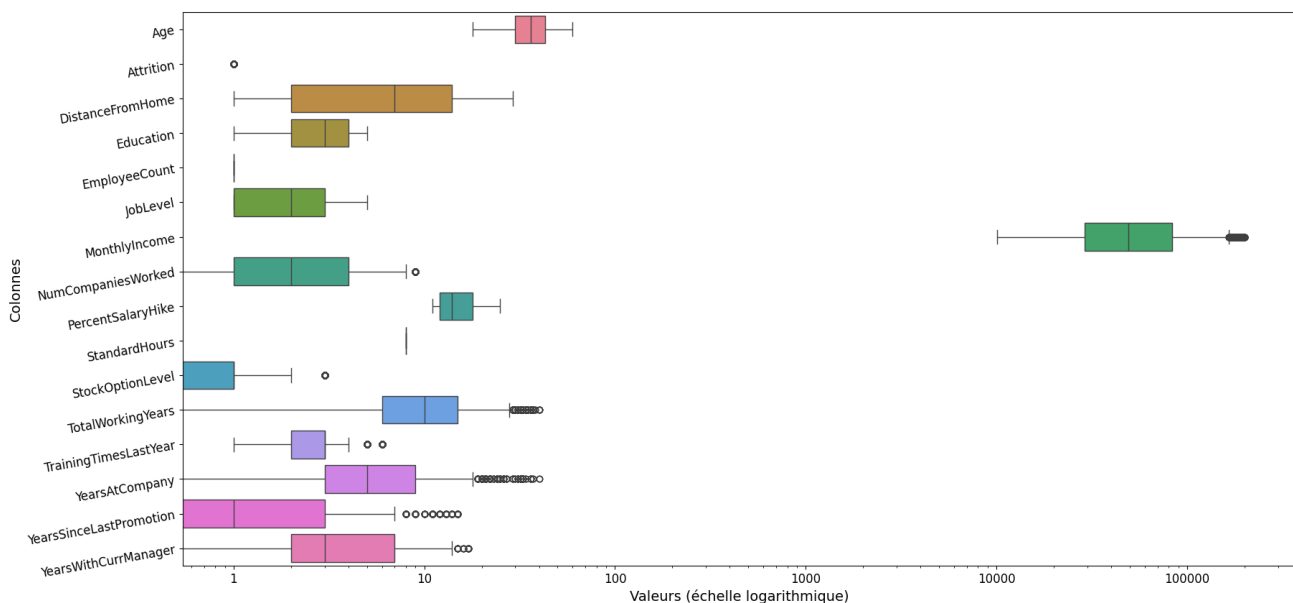


Figure 4 : Analyse Box Plot General Data

4. Informations sur le temps de travail (Work Info Data)

L'analyse des heures travaillées révèle :

- Une **moyenne d'heures travaillées par jour de 7,70 heures**, avec une variabilité modérée (écart-type de **1,34**).
- Le **nombre total d'heures travaillées** par employé varie entre **1348 et 2723 heures**, avec une moyenne de **1821 heures**.
- Les employés ont travaillé en moyenne **236 jours par an**, avec une variation faible.

	mean_worked_hours	total_worked_hours	worked_days
count	4410.000000	4410.000000	4410.000000
mean	7.700792	1821.276283	236.265306
std	1.340218	331.361528	5.503779
min	5.950504	1348.803056	225.000000
25%	6.673333	1563.572500	232.000000
50%	7.406761	1745.782222	236.000000
75%	8.368703	1967.295347	241.000000
max	11.030960	2723.378056	248.000000

Figure 5 : Analyse Descriptive Work Info

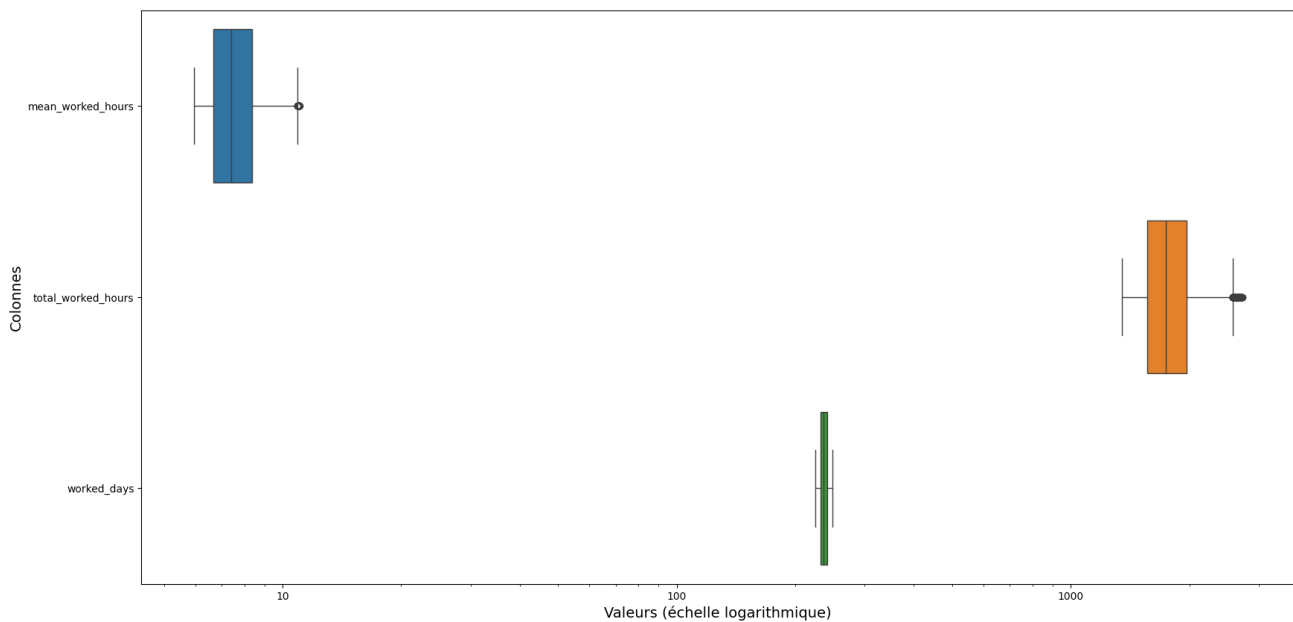


Figure 6 : Analyse Box Plot Work Info

E) Analyse des Relations

L'analyse du pairplot montre qu'aucune relation forte ne se dégage entre la plupart des variables. En effet, les nuages de points sont largement dispersés, ne laissant apparaître aucune tendance linéaire évidente. De plus, les relations entre les variables semblent faibles, voire inexistantes. Enfin, la diversité des distributions, avec certaines variables prenant des valeurs très variées ou de nature catégorielle, limite les possibilités de corrélations directes.

L'absence de corrélations fortes entre la majorité des variables est un point positif dans notre contexte, car cela signifie que :

- Indépendance des variables : Aucune variable unique ne domine l'explication d'une autre, ce qui montre que les différents facteurs influencent les employés de manière variée et équilibrée.
- Pas de biais évident : Il n'y a pas de facteur unique qui dicterait les comportements des employés.
- Une meilleure capacité de généralisation : Si nous utilisons ce dataset pour du machine learning (ex. prédire la satisfaction ou la rétention des employés), le modèle ne risque pas d'être biaisé par une relation trop dominante.
- Des décisions plus flexibles : L'entreprise peut prendre des décisions plus nuancées, sans être restreinte par une relation trop forte entre certaines variables (par exemple, elle ne doit pas nécessairement corréler systématiquement l'expérience au salaire).

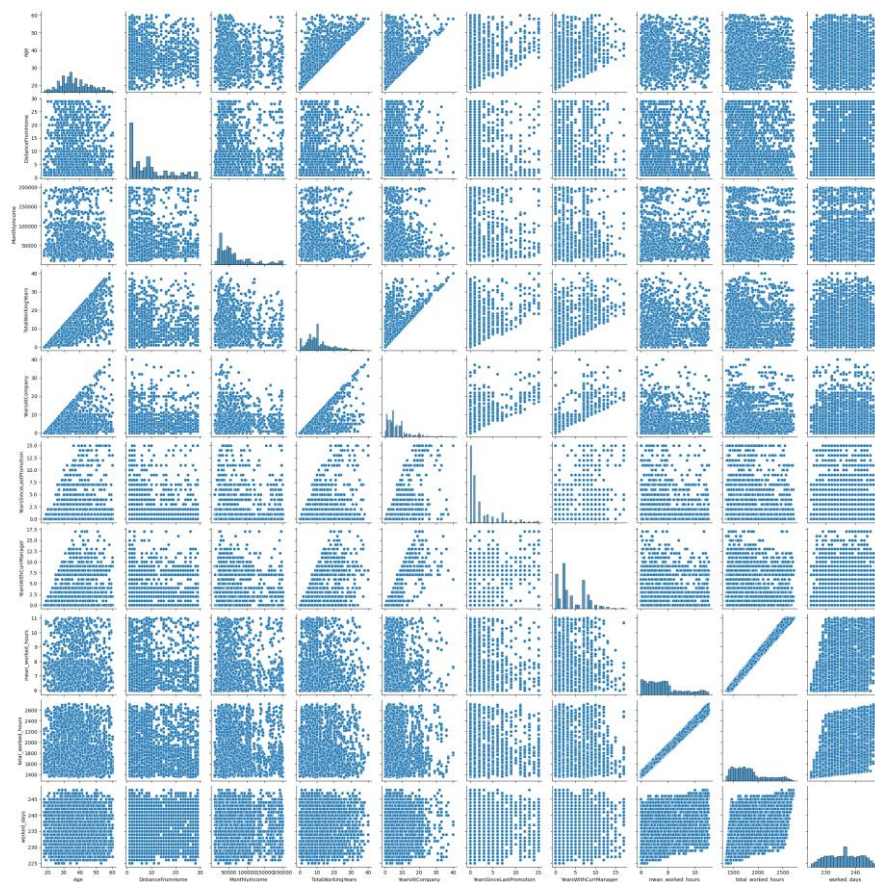


Figure 7 : Analyse PairPlot

Bien que la heatmap montre quelques corrélations modérées entre certaines variables (ex. TotalWorkingYears et âge), elles restent raisonnables et logiques. Aucune corrélation excessive ne vient fausser l'analyse, ce qui permet une évaluation plus objective des données.

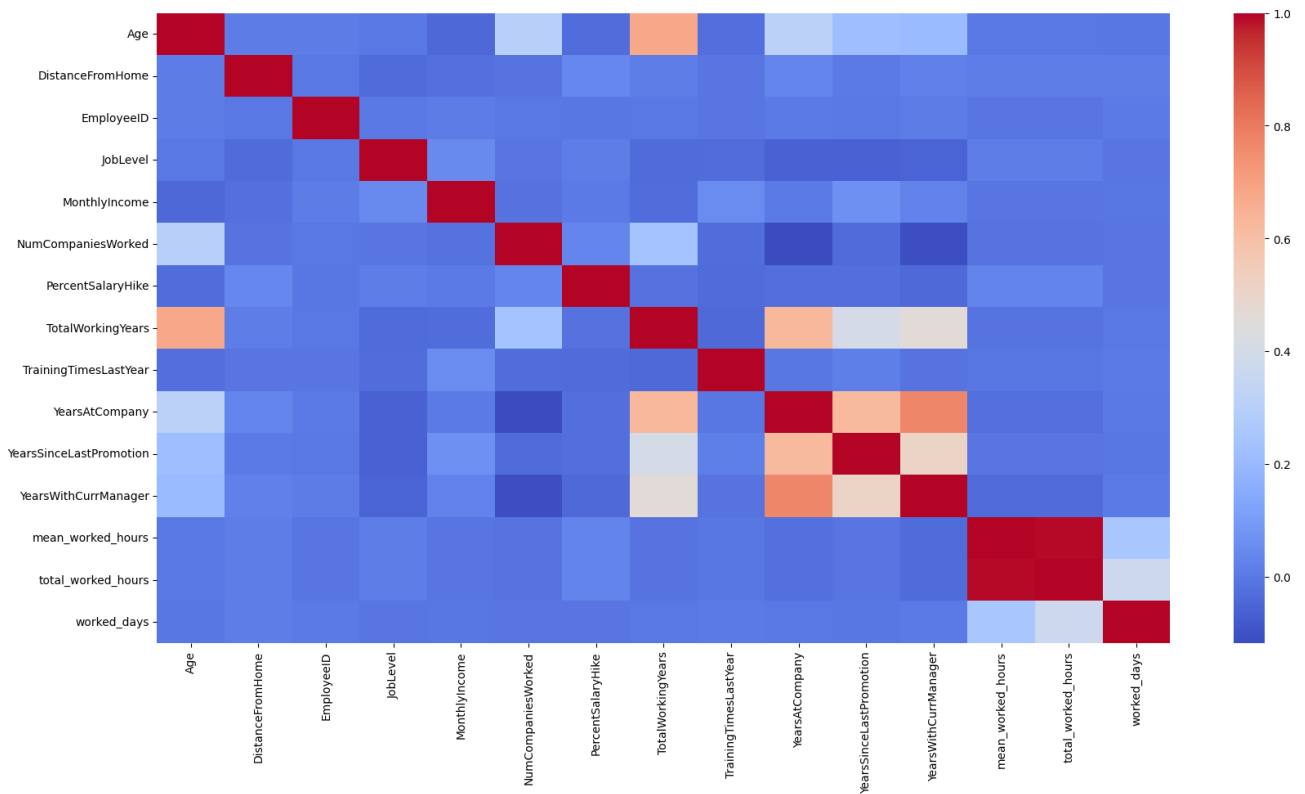


Figure 8 : Analyse HeatMap

V. Transformation des Données

Plusieurs étapes clés ont été réalisées pour optimiser la qualité des données et améliorer les performances des modèles prédictifs.

A) Création de nouvelles variables pertinentes

Une colonne `worked_hours` a été ajoutée, représentant le nombre total d'heures travaillées par jour pour chaque employé. Cette variable a permis de dériver trois indicateurs clés :

- Moyenne des heures travaillées par jour par employé.
- Nombre total d'heures travaillées par employé.
- Nombre de jours où l'employé a travaillé.

L'ajout de ces variables vise à enrichir le jeu de données en fournissant des informations supplémentaires qui peuvent améliorer la précision des modèles prédictifs. Selon le Conservatoire national des arts et métiers, la création de nouvelles variables permet de réduire l'information en un nombre de composantes plus limité que le nombre initial de variables, facilitant ainsi l'analyse et l'interprétation des données

B) Encodage des variables catégorielles (One-hot encoding)

Les variables catégorielles ont été transformées en variables numériques à l'aide de l'encodage one-hot. Cette technique consiste à créer des colonnes binaires pour chaque catégorie, indiquant la présence (1) ou l'absence (0) de cette catégorie pour chaque observation. Cette méthode est essentielle car de nombreux algorithmes d'apprentissage automatique ne peuvent pas traiter directement des données catégorielles. L'encodage one-hot permet ainsi de convertir ces variables en un format numérique compréhensible par les algorithmes, améliorant ainsi leur performance

C) Standardisation des données

La standardisation des données a été effectuée pour assurer que toutes les variables aient une échelle comparable. Cette étape est cruciale car de nombreux algorithmes de machine learning, tels que les méthodes de régression ou les réseaux de neurones, sont sensibles aux échelles des variables. En standardisant les données, on s'assure que chaque variable contribue de manière équitable à la modélisation, évitant ainsi que certaines variables dominent en raison de leur amplitude.

Ces techniques ont été appliquées en tenant compte des spécificités de notre jeu de données, afin d'assurer leur pertinence et leur efficacité dans le cadre de notre projet.

D) Rééquilibrage du Dataset pour une Meilleure Précision

Notre dataset contient un déséquilibre dans la répartition des classes de la variable cible (Attrition), cela peut entraîner un biais dans l'entraînement des modèles. Si une classe (True ou False) est surreprésentée, le modèle risque de privilégier cette classe et de mal généraliser sur la classe minoritaire. Pour résoudre ce problème, on peut appliquer une stratégie de sous-échantillonnage (undersampling), qui consiste à réduire la taille de la classe majoritaire pour l'aligner avec celle de la classe minoritaire. Cette méthode permet de garantir que le modèle apprend de manière équilibrée et évite de favoriser une classe au détriment de l'autre.

MODELISATION

I. METHODES DE VALIDATION DES MODELES

A) Validation Croisée

Pour s'assurer de la performance de nos modèles, nous avons systématiquement appliqué la validation croisée. Le principe est simple : nous divisons aléatoirement notre jeu de données en k sous-ensembles de taille égale (pour ce projet nous avons choisi $k=5$). Tour à tour, chaque sous-ensemble sert de jeu de test tandis que les $k-1$ autres sous-ensembles constituent le jeu d'entraînement. Cette approche permet d'obtenir une estimation plus robuste de la performance du modèle en réduisant la variance liée à un découpage unique des données. De plus, elle aide à détecter d'éventuels problèmes de surapprentissage en comparant les performances sur les différents plis. La moyenne des performances obtenues sur chaque pli fournit alors une mesure plus fiable de la capacité de généralisation du modèle.

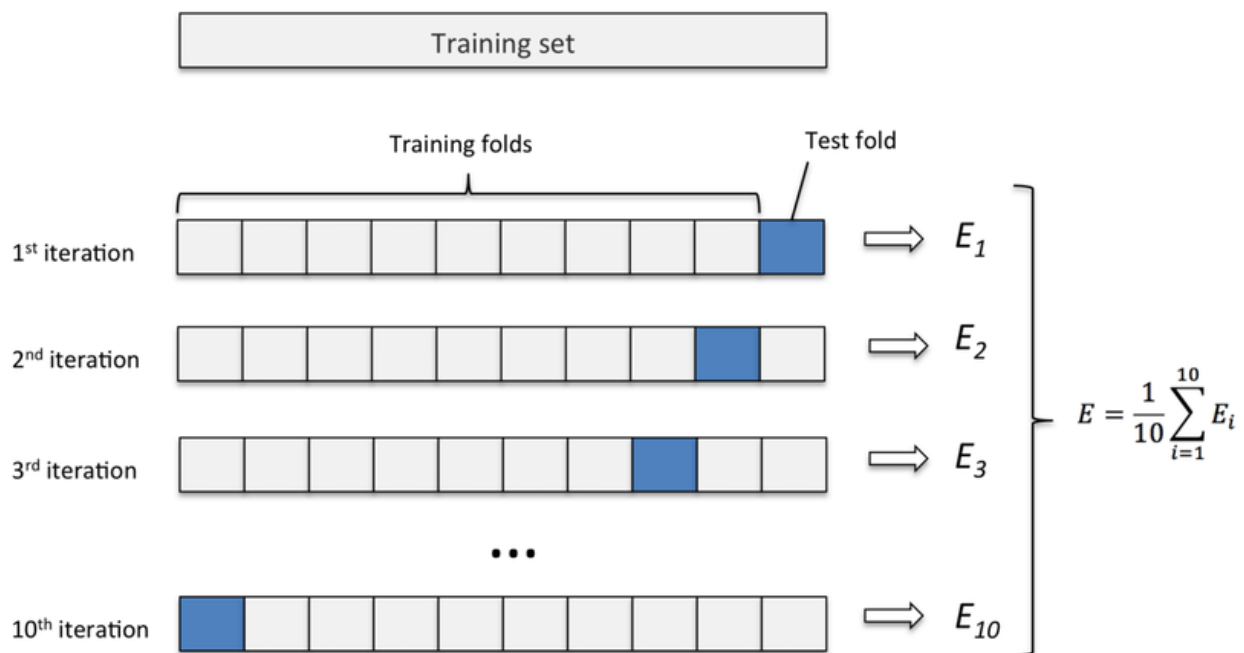


Figure 9: Principe de la validation croisée (exemple à 10 plis)

Source du schéma : Introduction to Support Vector Machines and Kernel Methods - Scientific Figure on ResearchGate. Available from: [Lien vers l'article](#) [accessed 5 Feb 2025]

Les résultats sont mesurés en faisant la moyenne des valeurs de chaque métrique pour les différents plis et en calculant l'écart-type.

B) Matrice de confusion

Pour nous aider à comprendre le type d'erreurs faites par notre modèle, nous avons utilisé des matrices de confusion.

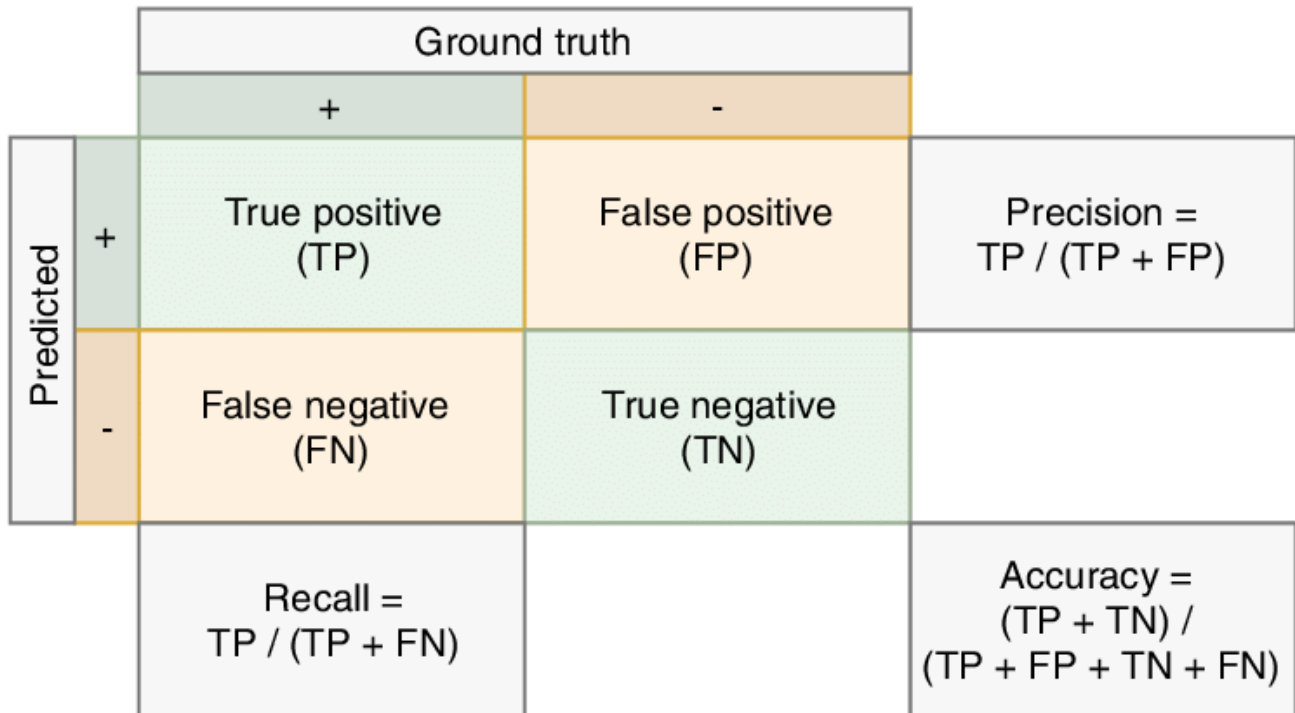


Figure 10: Matrice de confusion et méthode de calcul des métriques associées

Source du schéma : A cloud detection algorithm for satellite imagery based on deep learning - Scientific Figure on ResearchGate. Available from: [Lien vers l'article](#) [accessed 5 Feb 2025]

Dans la suite de ce rapport nous utiliserons les abréviations suivantes pour désigner les différentes parties de cette matrice de confusion :

VP (TP): Vrais positifs
VN (TN): Vrais négatifs
FP: Faux positifs
FN: Faux négatifs

C) Métriques

1) Accuracy

L'accuracy (précision globale) représente le pourcentage de prédictions correctes par rapport au nombre total de prédictions. Cette métrique, bien qu'intuitive, peut être trompeuse dans le cas de données déséquilibrées. Par exemple, pour un jeu de données contenant 95% de cas négatifs et 5% de cas positifs, un modèle qui prédirait systématiquement la classe majoritaire obtiendrait une accuracy de 95% sans pour autant être pertinent.

$$Accuracy = \frac{VP + VN}{VP + FP + VN + FN}$$

2) Précision

La précision mesure la proportion de vrais positifs parmi tous les cas prédits comme positifs. En d'autres termes, elle indique la capacité du modèle à éviter les faux positifs. Cette métrique est particulièrement importante dans les situations où les faux positifs sont coûteux, comme dans le diagnostic médical où l'on veut éviter de diagnostiquer à tort une maladie chez un patient sain.

$$Precision = \frac{VP}{VP + FP}$$

3) Rappel (Recall)

Le rappel (ou sensibilité) mesure la proportion de vrais positifs correctement identifiés parmi tous les cas réellement positifs. Il indique la capacité du modèle à détecter tous les cas positifs. Cette métrique est cruciale dans les situations où manquer un cas positif peut avoir des conséquences graves, comme dans la détection de fraudes bancaires où l'on souhaite identifier toutes les transactions frauduleuses.

$$Recall = \frac{VP}{VP + FN}$$

4) F1-Score

Le F1-score représente la moyenne harmonique entre la précision et le rappel. Cette métrique permet de trouver un équilibre entre la précision et le rappel, particulièrement utile lorsqu'il est nécessaire d'optimiser les deux métriques simultanément. Le F1-score est particulièrement pertinent dans les cas où les classes sont déséquilibrées, car il prend en compte à la fois les faux positifs et les faux négatifs.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

5) Courbe ROC

L'AUC (Area Under the ROC Curve) mesure la capacité du modèle à distinguer les classes à travers différents seuils de classification. Cette métrique représente la probabilité qu'un cas positif aléatoire soit classé avant un cas négatif aléatoire. Un AUC de 0.5 correspond à un classement aléatoire, tandis qu'un AUC de 1.0 indique une classification parfaite. Cette métrique est particulièrement utile car elle est insensible au déséquilibre des classes et permet de comparer différents modèles indépendamment du seuil de classification choisi.

II. IDENTIFICATION DES MODELES A TESTER

Afin de choisir les modèles les plus adaptés à notre problème de classification, nous avons pris exemple sur les modèles utilisés dans le workshop.

A) Régression Logistique

Le premier modèle que nous avons testé est un modèle de Régression Logistique inspiré de la méthode utilisée dans le workshop. La régression logistique permet d'estimer la probabilité d'appartenance à une classe en modélisant la relation entre un ensemble de variables explicatives et une variable de sortie binaire. Cette méthode est particulièrement adaptée aux problèmes de classification binaire où l'on cherche à prédire une variable catégorielle à deux modalités. Elle utilise une fonction sigmoïde pour transformer la combinaison linéaire des variables d'entrée en une probabilité comprise entre 0 et 1.

Dans notre exemple, c'est l'attrition que nous souhaitons prédire. L'attrition ne possède que 2 valeurs possibles donc la régression logistique semble adaptée à notre problème.

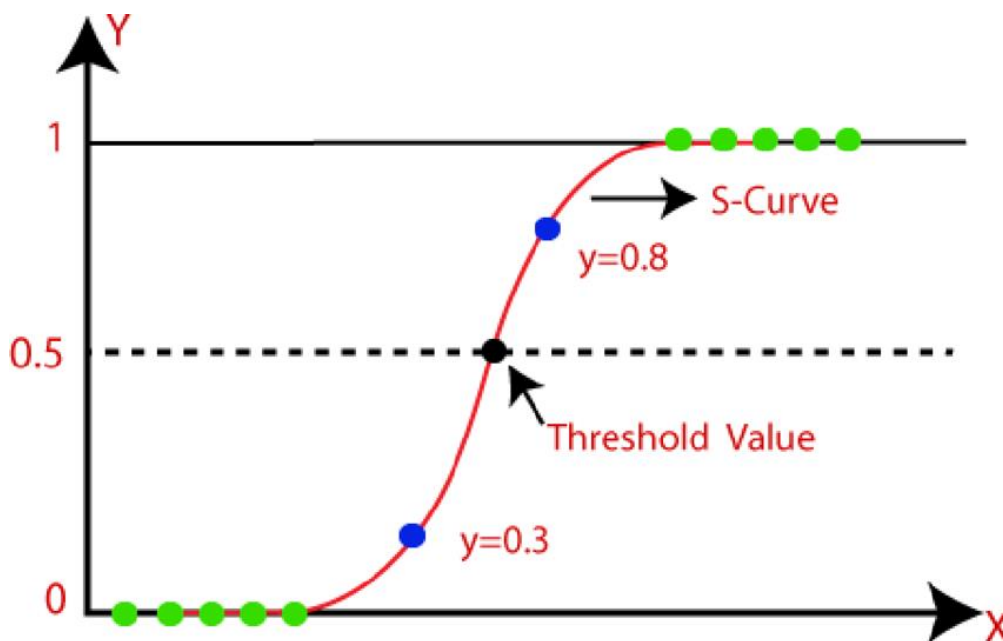


Figure 11: Classification avec la régression logistique

Source du schéma : An Intelligent System for Prediction of COVID-19 Case using Machine Learning Framework- Logistic Regression - Scientific Figure on ResearchGate. Available from: [Lien vers l'article](#) [accessed 5 Feb 2025]

Voici la formule correspondant à la régression logistique :

$$f(z) = \frac{1}{1 + e^{-z}}$$

Avec z une fonction multi-linéaire de la forme $z = b + \sum_{i=1}^n a_i x_i$ avec a_i le coefficient associé à la variable x_i et n le nombre de variables dans le dataset.

B) Support Vector Classifier (SVC)

Le Support Vector Classifier (SVC) est un algorithme de classification qui cherche à trouver l'hyperplan optimal permettant de séparer les différentes classes dans l'espace des caractéristiques. Son principe fondamental repose sur la maximisation de la marge, c'est-à-dire la distance entre l'hyperplan séparateur et les points les plus proches de chaque classe, appelés vecteurs de support. Pour gérer les cas non linéairement séparables, le SVC utilise le "kernel trick" qui permet de projeter les données dans un espace de dimension supérieure où elles deviennent linéairement séparables. Plusieurs types de noyaux sont disponibles (linéaire, polynomial, RBF, sigmoïde), chacun adapté à différents types de problèmes. Le choix des hyperparamètres, notamment le paramètre C qui contrôle la tolérance aux erreurs et les paramètres spécifiques au noyau choisi, est crucial pour optimiser les performances du modèle et éviter le surapprentissage.

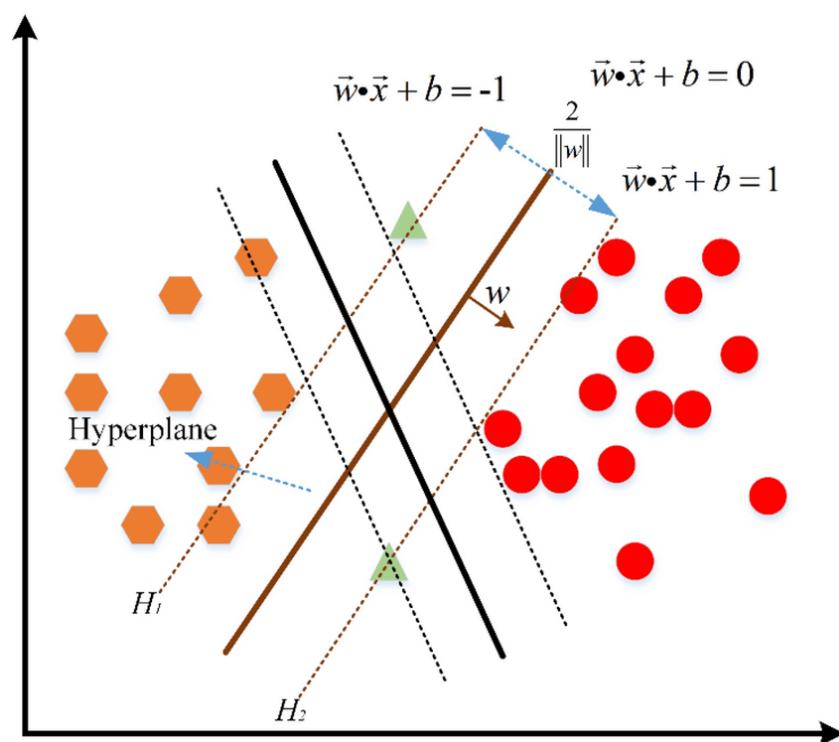


Figure 12: Schéma de la classification avec un SVC

Source du schéma : Rock mass classification prediction model using heuristic algorithms and support vector machines: a case study of Chambishi copper mine - Scientific Figure on ResearchGate. Available from: [Lien vers l'article](#) [accessed 5 Feb 2025]

C) Arbres de décision

Un arbre de décision est un modèle d'apprentissage supervisé qui fonctionne en divisant séquentiellement les données en sous-ensembles de plus en plus homogènes. À chaque nœud de l'arbre, l'algorithme sélectionne la caractéristique et le seuil qui permettent la meilleure séparation des données selon un critère d'impureté (comme l'indice de Gini ou l'entropie). Ce processus se répète récursivement sur chaque sous-ensemble jusqu'à atteindre un critère d'arrêt, comme une profondeur maximale ou un nombre minimal d'échantillons par feuille. Les prédictions sont alors faites en parcourant l'arbre de la racine jusqu'à une feuille, où la classe majoritaire (pour la classification) ou la moyenne des valeurs (pour la régression) détermine la sortie.

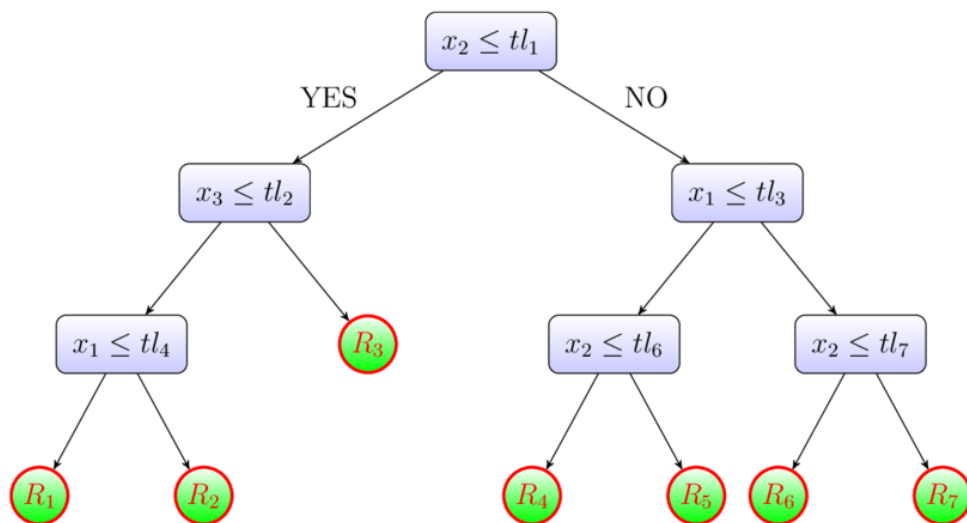


Figure 13 : Schéma d'un arbre de décision

Source du schéma : Drivers of economic and financial integration: A machine learning approach - Scientific Figure on ResearchGate. Available from: [Lien vers l'article](#) [accessed 5 Feb 2025]

D) Forêts aléatoires

Bien que simples à interpréter et à visualiser, les arbres de décision peuvent souffrir de surapprentissage, ce qui nécessite souvent l'utilisation de techniques de régularisation ou leur intégration dans des ensembles plus robustes comme les forêts aléatoires. Les forêts aléatoires (Random Forests) sont des ensembles d'arbres de décision entraînés de manière indépendante selon le principe du bagging (Bootstrap Aggregating). Chaque arbre est construit sur un sous-ensemble aléatoire des données d'origine, tiré avec remise, et utilise à chaque division un sous-ensemble aléatoire des caractéristiques disponibles (typiquement \sqrt{p} caractéristiques où p est le nombre total de caractéristiques). La prédiction finale est obtenue par vote majoritaire (pour la classification) ou par moyenne (pour la régression) des prédictions de tous les arbres. Cette approche permet de réduire significativement le surapprentissage inhérent aux arbres individuels tout en conservant leur capacité à capturer des relations non linéaires complexes. L'introduction de l'aléatoire à deux niveaux (échantillonnage des données et des caractéristiques) garantit une faible corrélation entre les arbres, ce qui améliore la robustesse et la généralisation du modèle final. Les forêts aléatoires offrent également une mesure native de l'importance des caractéristiques, facilitant ainsi l'interprétation du modèle.

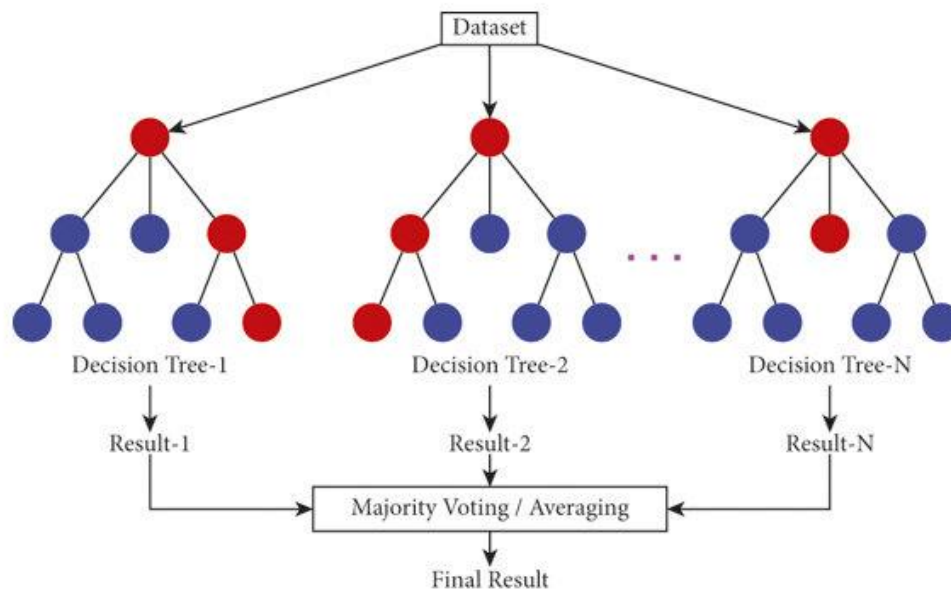


Figure 14: Principe des forêts aléatoires

Source du schéma : Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques - Scientific Figure on ResearchGate. Available from: [Lien vers l'article](#) [accessed 5 Feb 2025]

III. RESULTATS

Pour comparer nos résultats sur les différents modèles, nous avons tracé la courbe ROC correspondant à chacun des 4 modèles choisis afin d'évaluer leur capacité à distinguer les classes positives et négatives. La courbe ROC représente le taux de vrais positifs en fonction du taux de faux positifs pour plusieurs seuils de classification, permettant ainsi d'analyser et de comparer leurs performances.

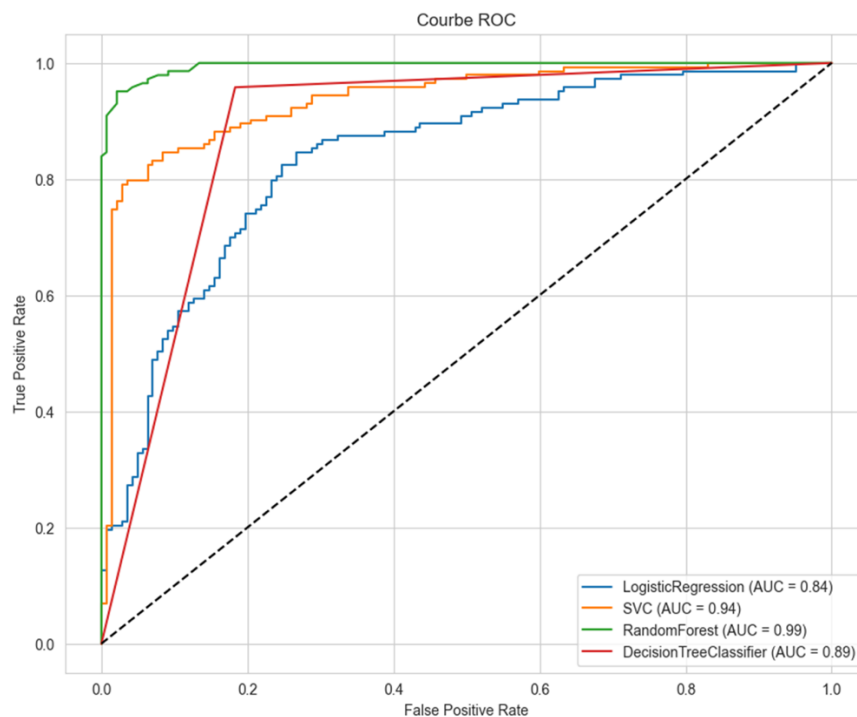


Figure 15 : Courbes ROC pour nos 4 modèles

Interprétation des résultats :

- Logistic Regression (AUC = 0.84) : Ce modèle présente une performance correcte avec une aire sous la courbe (AUC) de 0.84, indiquant une bonne séparation entre les classes mais moins efficace que d'autres modèles.
- SVC (AUC = 0.94) : Le Support Vector Classifier (SVC) offre de meilleures performances avec un AUC de 0.94, ce qui signifie qu'il discrimine bien les classes.
- Random Forest (AUC = 0.99) : Ce modèle obtient une excellente performance avec un AUC de 0.99, indiquant qu'il distingue presque parfaitement les classes positives et négatives.
- Decision Tree Classifier (AUC = 0.89) : L'arbre de décision montre une bonne performance avec un AUC de 0.89, bien que légèrement inférieur à celui du SVC et du Random Forest.

Le modèle Random Forest se distingue comme étant le plus performant parmi les quatre modèles testés, tandis que la régression logistique est la moins efficace en termes de discrimination des classes. En fonction des besoins (précision, interprétabilité, complexité du modèle), il conviendra de choisir le modèle le plus adapté.

IV. OPTIMISATION DES HYPERPARAMETRES

L'optimisation des hyperparamètres est une étape essentielle dans le développement de modèles d'apprentissage automatique, car elle permet d'améliorer leurs performances en ajustant les paramètres influençant leur comportement. Dans ce projet, nous avons utilisé la recherche par grille (GridSearchCV) pour identifier les meilleures configurations des algorithmes Support Vector Classifier (SVC) et Random Forest, en testant systématiquement différentes combinaisons de paramètres.

A) Méthodologie

La validation croisée à 5 plis a été employée pour évaluer la robustesse des modèles tout en réduisant le risque de sur-ajustement. Plusieurs métriques ont été prises en compte, notamment l'accuracy, le score F1, la précision, le rappel et l'aire sous la courbe ROC (AUC-ROC), afin d'obtenir une évaluation complète des performances. Le critère de sélection du meilleur estimateur pour la validation croisée a été basé sur l'accuracy.

Les paramètres testés pour chaque modèle étaient les suivants :

- **Support Vector Machine (SVC)** : les hyperparamètres ajustés incluaient le coefficient de régularisation (C), le type de noyau (*kernel*), le paramètre de lissage (*gamma*), ainsi que la prise en compte d'un équilibrage des classes (*class_weight*).
- **Random Forest** : l'optimisation portait sur le nombre d'arbres ($n_estimators$), la profondeur maximale (*max_depth*), la stratégie de sélection des caractéristiques (*max_features*), et le critère de division des nœuds (*min_samples_split*).

B) Résultats de l'optimisation

Après l'exploration des différentes configurations, les meilleurs paramètres retenus sont :

- **Random Forest** :
 - Poids des classes : équilibré
 - Profondeur maximale : 20
 - Nombre d'arbres : 300
 - Nombre minimal d'échantillons pour diviser un nœud : 2
 - Sélection des caractéristiques : *log2*
- **SVC** :
 - Coefficient de régularisation (C) : 10
 - Noyau : *rbf*
 - Paramètre gamma : *scale*
 - Poids des classes : non équilibré

En termes de performances, la Random Forest a obtenu un score de validation croisée de 93,1% pour l'accuracy, surpassant légèrement le SVC, qui a atteint 91,7%. Ces résultats démontrent une meilleure capacité de généralisation du modèle Random Forest sur cet ensemble de données.

C) Choix du modèle

D'après les résultats précédents, la forêt aléatoire semble avoir de meilleurs résultats sur nos données. En effet les moyennes pour chaque métrique sont plus élevées pour la forêt aléatoire et les écarts types sont plus faibles.

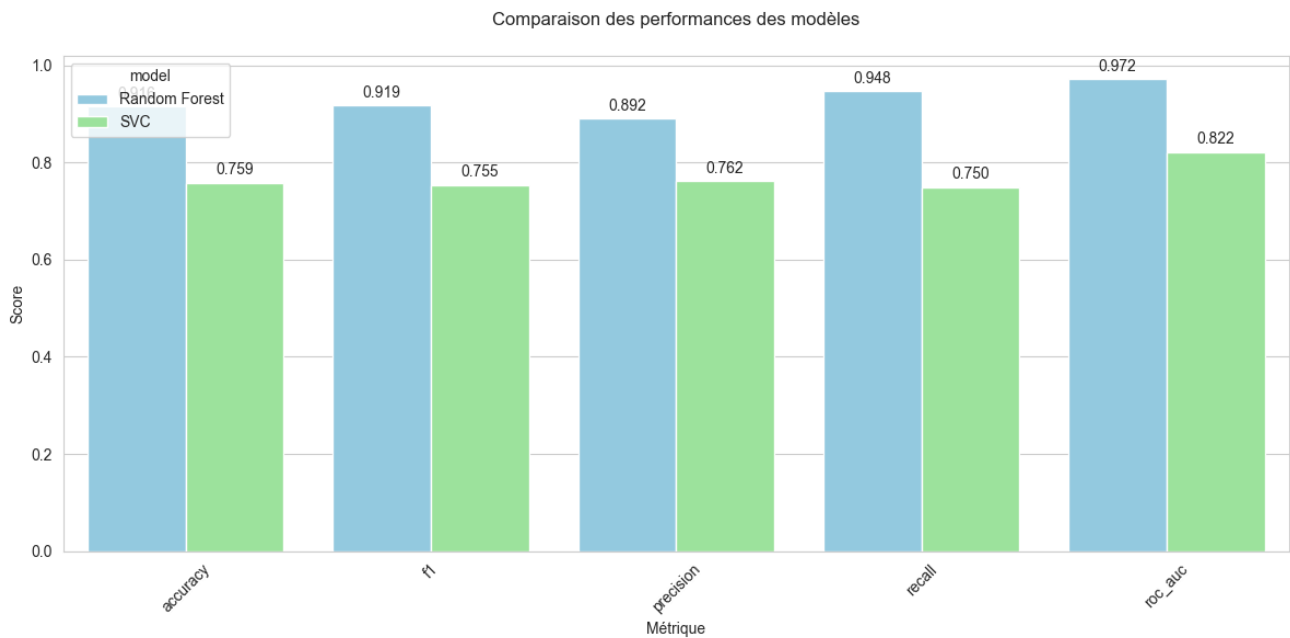


Figure 16: Comparaison des performances des modèles optimisés

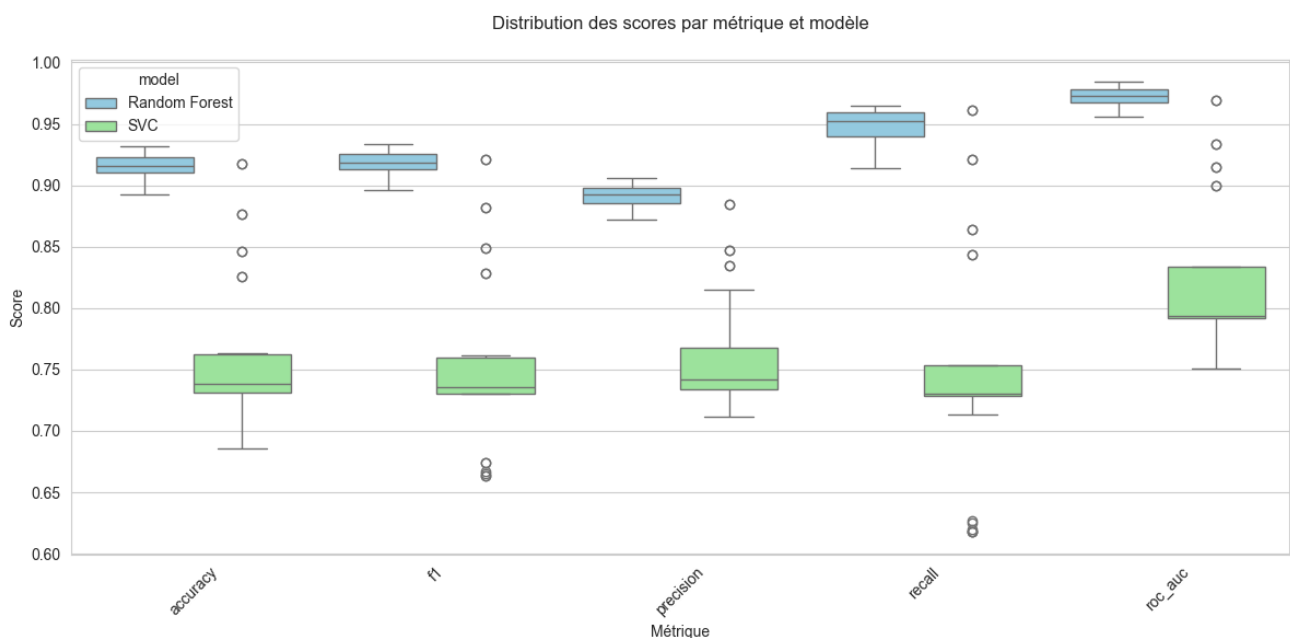


Figure 17: Distribution des scores par métrique et par modèle

FACTEURS AYANT LE PLUS D'IMPACT

Notre client, HumanForYou, nous a demandé d'identifier les facteurs ayant le plus d'impact sur le taux d'attrition. Nous avons donc utilisé plusieurs méthodes pour identifier ces facteurs à partir des données :

- **VarianceThreshold** : permet d'éliminer les caractéristiques dont la variance est inférieure à un seuil prédéfini. Cette méthode est particulièrement utile pour supprimer les variables quasi-constantes qui apportent peu d'information au modèle. Elle constitue souvent une première étape de filtrage des variables peu informatives. Nous avons choisi un seuil de variance de 0.5 pour que le facteur soit considéré comme important.
- **SelectFromModel avec RandomForest** : permet de sélectionner les variables les plus importantes en se basant sur les scores d'importance des caractéristiques calculés par la forêt aléatoire. Cette méthode exploite la capacité des forêts aléatoires à quantifier naturellement l'importance de chaque variable à travers la diminution moyenne de l'impureté (mean decrease impurity) ou la diminution moyenne de la précision (mean decrease accuracy).
 - **Cela revient à utiliser les coefficients du meilleur modèle.** Cela permet d'identifier les variables les plus influentes en analysant directement les coefficients ou les scores d'importance du modèle final retenu. Pour les modèles linéaires, cela correspond à l'amplitude des coefficients, tandis que pour les modèles plus complexes comme les forêts aléatoires, cela peut être basé sur les scores d'importance des caractéristiques intégrés au modèle.
- **RFE (Recursive Feature Elimination)** : permet de sélectionner les caractéristiques de manière itérative en entraînant le modèle, en calculant l'importance des caractéristiques, et en éliminant récursivement les moins importantes. À chaque itération, les variables les moins pertinentes sont éliminées jusqu'à atteindre le nombre désiré de caractéristiques. Cette approche est plus sophistiquée car elle prend en compte les interactions entre variables lors du processus de sélection.

	VARIANCE THRESHOLD	SELECT FROM MODEL	RFE
0	Age	Age	Age
1	DistanceFromHome	DistanceFromHome	DistanceFromHome
2	JobLevel	JobLevel	JobLevel
3	MonthlyIncome	MonthlyIncome	MonthlyIncome
4	NumCompaniesWorked	NumCompaniesWorked	NumCompaniesWorked
5	PercentSalaryHike	PercentSalaryHike	PercentSalaryHike
6	TotalWorkingYears	TotalWorkingYears	TotalWorkingYears
7	TrainingTimesLastYear	TrainingTimesLastYear	TrainingTimesLastYear
8	YearsAtCompany	YearsAtCompany	YearsAtCompany
9	YearsSinceLastPromotion	YearsSinceLastPromotion	YearsSinceLastPromotion
10	YearsWithCurrManager	YearsWithCurrManager	YearsWithCurrManager
11	mean_worked_hours	mean_worked_hours	mean_worked_hours
12	total_worked_hours	total_worked_hours	total_worked_hours
13	worked_days	worked_days	worked_days
14	None	MaritalStatus_Single	MaritalStatus_Single
15	None	EnvironmentSatisfaction_Poor	None

Nous avons aussi classé ces facteurs en utilisant les coefficients de chaque modèle que nous avons testé.

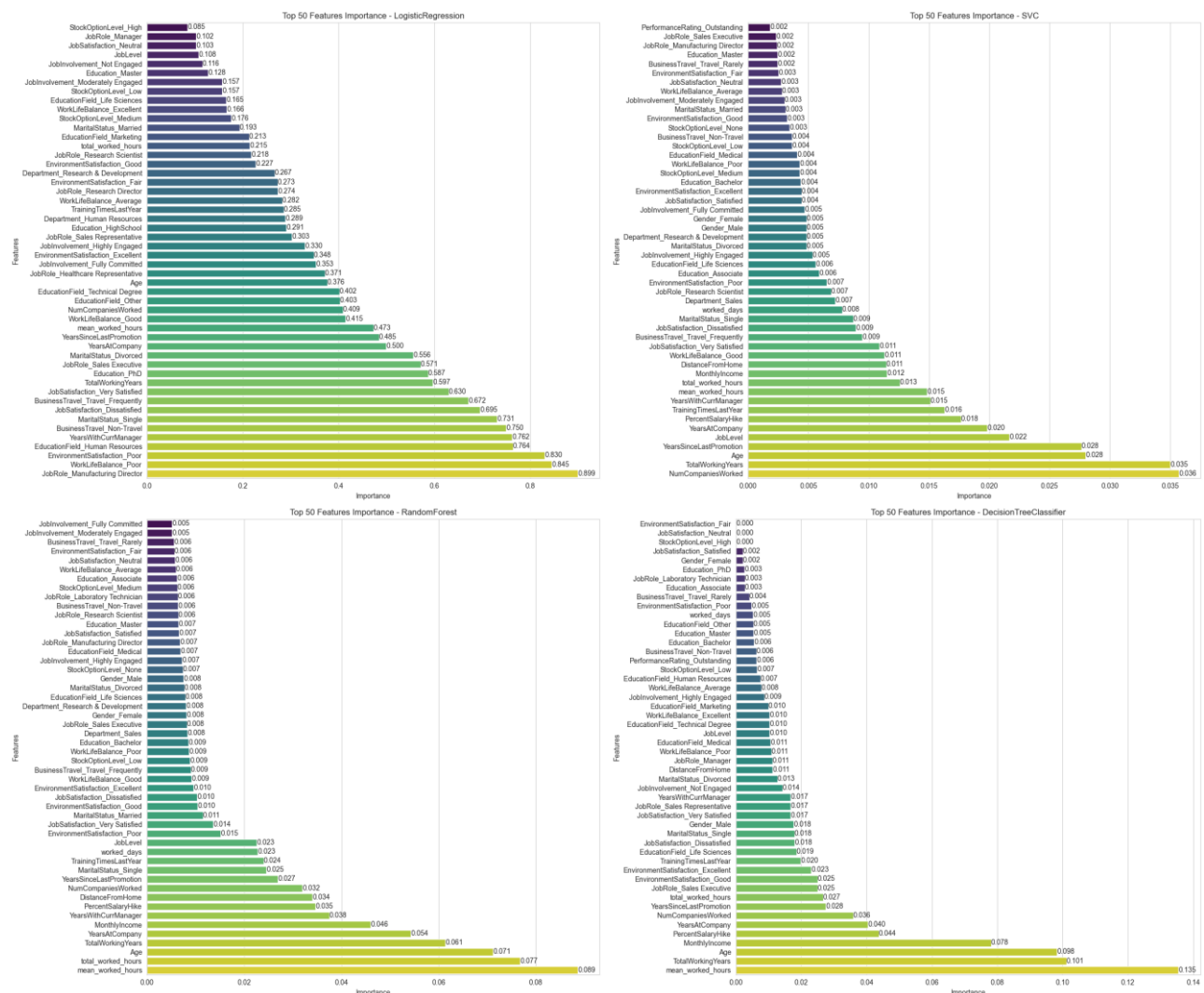


Figure 18: Comparaison des facteurs ayant le plus d'impact sur l'attrition selon les modèles

Malgré des variations selon les modèles et les méthodes, les 5 facteurs les plus importants semblent être :

- Age : Age du salarié en 2015.
- DistanceFromHome : Distance en km entre le logement de l'employé et l'entreprise.
- JobLevel : Niveau hiérarchique dans l'entreprise de 1 à 5
- NumCompaniesWorked : Nombre d'entreprises pour lequel le salarié a travaillé avant de rejoindre HumanForYou.
- Mean worked hours : Nombre moyen d'heures travaillées par jour par l'employé.

CONSEILS A L'ENTREPRISE

I. AGE DES EMPLOYES

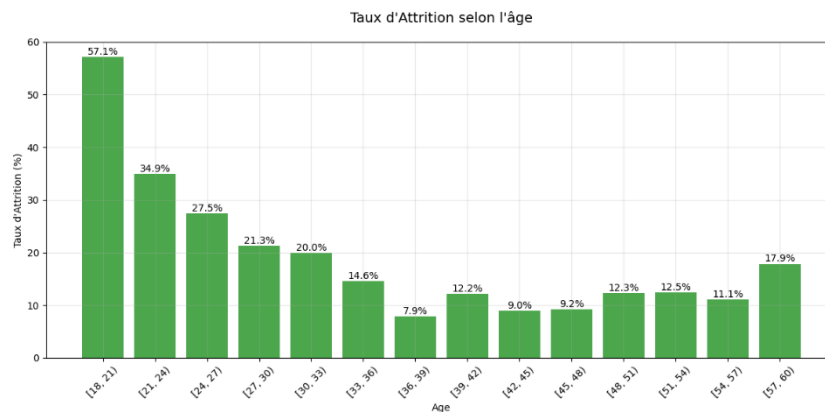


Figure 19: Attrition selon l'âge

D'après ce graphique, on constate que ce sont les plus jeunes (33 ans et moins) qui ont le plus tendance à partir de l'entreprise. Une piste pour résoudre ce problème serait de mettre en place des politiques de rétention spécifiques pour les jeunes employés, telles que des opportunités de développement de carrière, des formations, des avantages sociaux attractifs ou encore un meilleur équilibre entre vie professionnelle et personnelle. Cela pourrait les encourager à rester plus longtemps au sein de l'entreprise.

II. DISTANCE DOMICILE-TRAVAIL

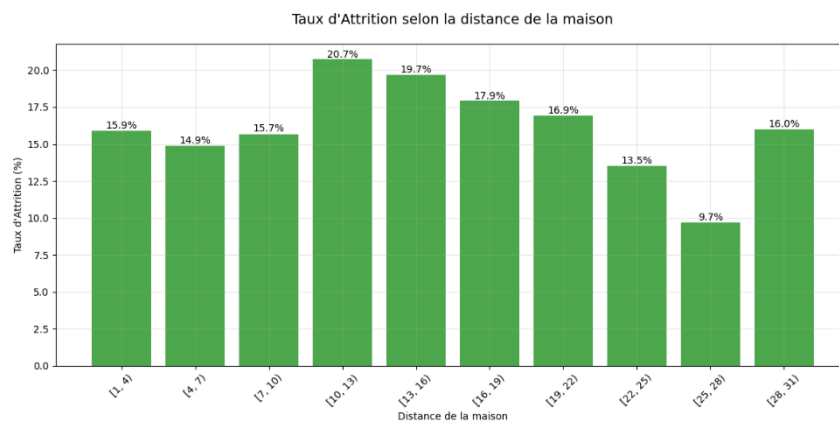


Figure 20: Attrition selon la distance domicile-travail

D'après ce graphique, il serait pertinent de mettre en place une aide au déménagement afin de faciliter l'installation des employés à proximité (moins de 10 km) de leur lieu de travail. Cette initiative pourrait contribuer à réduire les contraintes liées aux trajets quotidiens, diminuer la fatigue et améliorer la satisfaction des employés, ce qui pourrait ainsi favoriser leur engagement et leur rétention au sein de l'entreprise.

III. NIVEAU HIERARCHIQUE DANS L'ENTREPRISE

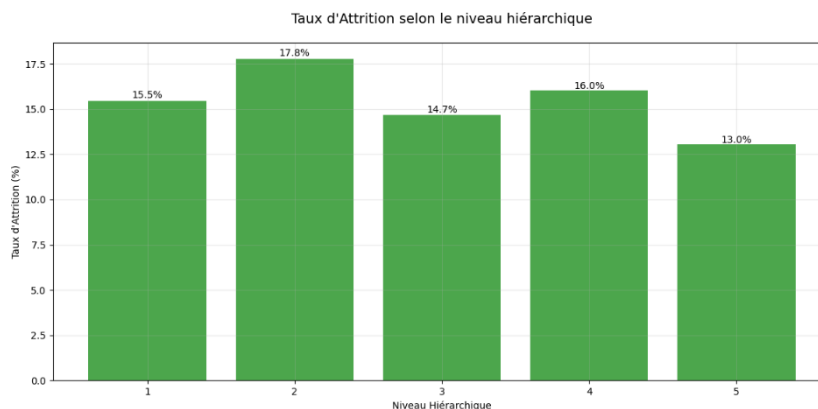


Figure 21: Attrition selon le niveau hiérarchique

D'après ce graphique, les personnes avec un niveau hiérarchique de 2 ont un taux d'attrition plus important. Il pourrait être pertinent d'analyser les raisons spécifiques de ce taux d'attrition élevé en menant des enquêtes internes ou des entretiens de sortie. En fonction des résultats, l'entreprise pourrait mettre en place des actions ciblées, telles qu'une meilleure reconnaissance du travail accompli, des opportunités de formation et d'évolution, une révision des conditions salariales ou encore un accompagnement managérial plus adapté afin d'améliorer la satisfaction et la rétention de ces employés.

IV. NOMBRE D'ENTREPRISES PRECEDENTES

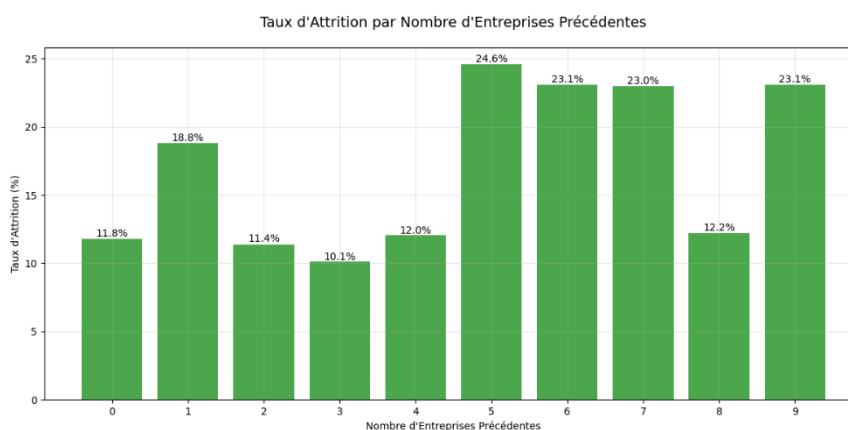


Figure 22: Attrition selon le nombre d'entreprises précédentes

D'après ce graphique, les personnes ayant travaillé dans exactement 1 entreprise ou celles ayant travaillé dans 5 entreprises ou plus sont celles qui ont le taux d'attrition le plus élevé. Pour atténuer ce phénomène, nous recommandons de mettre en place des stratégies adaptées à ces deux profils distincts. Pour les employés ayant travaillé dans exactement 1 entreprise, il pourrait être utile de proposer des perspectives d'évolution claires et des opportunités de développement professionnel afin d'éviter qu'ils ne ressentent le besoin de changer d'entreprise à nouveau. Pour ceux ayant travaillé dans 5 entreprises ou plus, il serait pertinent d'analyser les raisons de leur forte mobilité et de renforcer les dispositifs d'intégration, de mentorat et d'engagement à long terme, en mettant l'accent sur la culture d'entreprise et la fidélisation.

V. MOYENNE D'HEURES TRAVAILLEES PAR JOUR

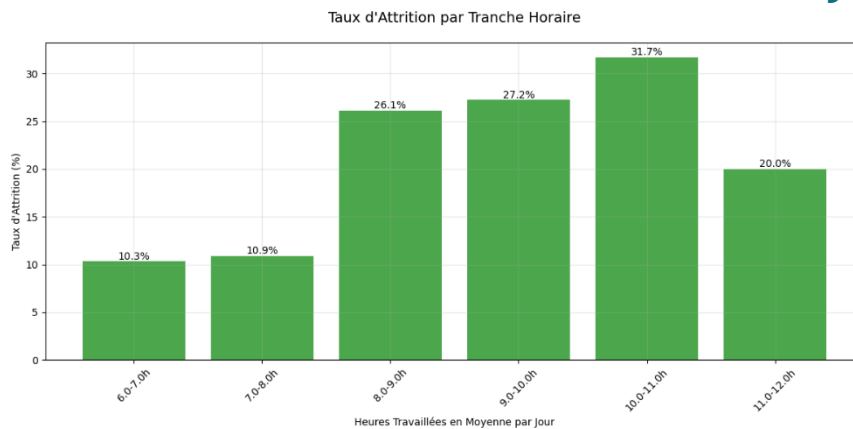


Figure 23: Attrition selon le nombre moyen d'heures travaillées par jour

D'après ce graphique, les personnes travaillant plus de 8h par jour ont le taux d'attrition le plus important. Pour régler ce problème, il est possible de mettre en place des mesures visant à améliorer l'équilibre entre vie professionnelle et vie personnelle. Cela pourrait inclure la promotion du télétravail, une meilleure gestion de la charge de travail, l'instauration de plages horaires plus flexibles ou encore l'encouragement à prendre des pauses et des congés. De plus, sensibiliser les managers à l'importance du bien-être des employés et optimiser la répartition des tâches pourraient contribuer à réduire la surcharge de travail et, par conséquent, le taux d'attrition.

BIBLIOGRAPHIE ANNOTÉE

- (1) Alduayj, S. S., & Rajpoot, K. (2018). Predicting Employee Attrition using Machine Learning. *2018 International Conference on Innovations in Information Technology (IIT)*, 93–98. [Lien vers l'article](#)

Cette étude analyse l'utilisation de modèles d'apprentissage automatique pour prédire l'attrition des employés, en se basant sur des données synthétiques fournies par IBM Watson. Elle a été déterminante dans le choix de nos modèles pour anticiper le départ des employés au sein de l'organisation.

- (2) Aschbacher, K. (n.d.). *The Uncertain Promise of HR Analytics: A Comprehensive Analysis of Data Utilised in Predictive Models for Employee Turnover*. [Lien vers l'article](#)

Cet article nous a aidés à comprendre les limites et incertitudes des modèles prédictifs utilisés en gestion des ressources humaines. Il souligne les défis liés à la qualité des données et aux biais algorithmiques, ce qui a nourri notre réflexion sur les limites et défis de l'IA appliquée au turnover.

- (3) *Enjeux, risques éthiques et cybersécurité de l'Intelligence Artificielle*. (n.d.). CentraleSupélec Exed. Retrieved 3 February 2025, from [Lien vers l'article](#)

Cet article souligne l'importance de l'éthique et de la cybersécurité dans le développement de l'intelligence artificielle, en abordant des problématiques telles que les biais algorithmiques, la transparence et la protection des données personnelles. Il a guidé notre réflexion sur les considérations éthiques à intégrer dans notre projet IA.

- (4) Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E. (2020). Predicting Employee Attrition Using Machine Learning Techniques. *Computers*, 9(4), Article 4. [Lien vers l'article](#)

Cet article explore l'application de techniques d'apprentissage automatique pour prédire l'attrition des employés, en mettant l'accent sur l'importance de l'analyse exploratoire des données (EDA) pour comprendre les facteurs influençant le départ des employés. Il nous a permis de mieux structurer notre approche EDA dans le cadre de notre projet.

- (5) Hendrycks, D. (2025). *Introduction to AI Safety, Ethics, and Society*. Taylor & Francis. [Lien vers l'article](#)

Ce livre a été une ressource clé dans l'élaboration de notre rapport sur l'éthique en IA. Il explore les risques liés à l'IA, notamment les enjeux de sécurité, de gouvernance et d'utilisation responsable, en fournissant une approche interdisciplinaire essentielle à notre réflexion.

- (6) Morelli, C., Fusai, G., & Zenti, R. (2024). Who and why will leave me? Utilizing Machine Learning-Based Models to Anticipate and Manage Employee Turnover. *SSRN Electronic Journal*. [Lien vers l'article](#)

Cet article a contribué à notre analyse des stratégies d'engagement des employés basées sur l'IA. Il met en évidence l'importance des recommandations personnalisées pour la rétention du personnel, notamment via l'optimisation des parcours de carrière.

- (7) Nosratabadi, S., Zahed, R. K., Ponkratov, V. V., & Kostyrin, E. V. (2022). Artificial Intelligence Models and Employee Lifecycle Management: A Systematic Literature Review. *Organizacija*, 55(3), 181–198. [Lien vers l'article](#)

Cette revue systématique nous a permis de mieux comprendre les modèles d'IA appliqués à la gestion du cycle de vie des employés. Elle a été particulièrement utile pour approfondir les approches de prédiction du turnover, telles que les réseaux de neurones et forêts aléatoires.

- (8) (PDF) *EMPLOYEE ATTRITION PREDICTION IN INDUSTRY USING MACHINE LEARNING TECHNIQUES*. (n.d.). Retrieved 3 February 2025, from [Lien vers l'article](#)

Cet article nous a aidé pour choisir et comprendre les modèles de machine learning utilisés dans notre projet. Il compare diverses méthodes de sélection de caractéristiques et de classification, permettant d'optimiser la précision des prédictions et de réduire les erreurs.

- (9) (PDF) *PREDICTION OF EMPLOYEE ATTRITION USING DATAMINING*. (n.d.). *ResearchGate*. [Lien vers l'article](#)

L'article explore diverses techniques de data mining pour prédire l'attrition des employés, offrant des perspectives précieuses sur les approches analytiques efficaces dans ce domaine.

- (10) Stuart Russell, Peter Norvig, Fabrice Popineau, Laurent Miclet, Claire Cadet. *Intelligence artificielle : une approche moderne* (4^e édition). Pearson France, 2021, 978-2-3260-0221-0. {hal-04245057}

Ce livre nous a été particulièrement utile pour traiter les enjeux éthiques de l'IA. Il propose une analyse approfondie des implications morales et sociétales des systèmes intelligents, apportant des éclairages essentiels à notre réflexion sur l'éthique en IA.

- (11) Yircof (2019). "Predicting Employee Turnover Using NLP". [Lien vers l'article](#)

Cette étude a été utilisée pour illustrer l'apport du traitement du langage naturel (NLP) dans la détection des signaux d'insatisfaction des employés. Elle met en avant l'analyse sémantique des échanges internes pour identifier les risques de départ.