



**HAL**  
open science

## L'utilisation du machine Learning pour favoriser la gestion proactive des ressources humaines : Etude de cas

Zakarya Laghzal, Lamya Temnati

### ► To cite this version:

Zakarya Laghzal, Lamya Temnati. L'utilisation du machine Learning pour favoriser la gestion proactive des ressources humaines : Etude de cas. *International Journal of Accounting, Finance, Auditing, Management and Economics*, 2022, 3, pp.547-563. 10.5281/zenodo.6582612 . hal-03787323

**HAL Id: hal-03787323**

**<https://hal.science/hal-03787323>**

Submitted on 24 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **L'utilisation du machine Learning pour favoriser la gestion proactive des ressources humaines : Etude de cas**

### **Using Machine Learning to Promote Proactive Human Resources Management: A Case Study**

**Zakarya LAGHZAL, (Doctorant chercheur)**

*LERSEM*

*École Nationale de Commerce et de Gestion  
Université Chouaib Doukkali El Jadida, Maroc*

**Lamya TEMNATI, (Enseignante chercheuse)**

*LERSEM*

*École Nationale de Commerce et de Gestion  
Université Chouaib Doukkali El Jadida, Maroc*

<b>Adresse de correspondance :</b>	ENCG El Jadida Université Choiaib Doukkali Maroc (El Jadida) 05233-94435
<b>Déclaration de divulgation :</b>	Les auteurs n'ont pas connaissance de quelconque financement qui pourrait affecter l'objectivité de cette étude.
<b>Conflit d'intérêts :</b>	Les auteurs ne signalent aucun conflit d'intérêts.
<b>Citer cet article</b>	LAGHZAL, Z., & TEMNATI, L. (2022). L'utilisation du machine Learning pour favoriser la gestion proactive des ressources humaines : Etude de cas. International Journal of Accounting, Finance, Auditing, Management and Economics, 3(3-2), 547-563. <a href="https://doi.org/10.5281/zenodo.6582612">https://doi.org/10.5281/zenodo.6582612</a>
<b>Licence</b>	<b>Cet article est publié en open Access sous licence CC BY-NC-ND</b>

*Received: May 09, 2022*

*Published online: May 31, 2022*

## L'utilisation Du Machine Learning Pour Favoriser La Gestion Proactive Des ressources Humaines : Étude De Cas

### Résumé :

Depuis la révolution industrielle, la fonction des ressources humaines (RH) a subi plusieurs changements, aujourd'hui elle est sollicitée d'adopter une approche à long terme et proactive qui sert à anticiper les besoins de l'entreprise et les problèmes susceptibles d'impacter sa productivité et sa performance afin de mettre en place des actions d'adaptation à long terme et être efficace dans sa démarche stratégique. Par ailleurs la machine Learning (ou autoapprentissage) est une technologie tendance depuis quelque temps, qui a connu une utilisation massive dans de nombreux domaines. Selon une étude menée par les décideurs IT issus de plus de 15 secteurs d'activités différents dans la Royaume uni, la France, l'Allemand et l'Espagne, 87% de l'échantillon ont implémenté cette technologie ou prévoient de le faire. Cette technologie a permis aux entreprises d'améliorer leur processus, augmenter leur compétitivité et aider à la prise de décision dans plusieurs domaines de gestion tels que la finance et le marketing.

Ce présent travail cherche à mettre en exergue le potentiel de cette technologie pour favoriser une gestion proactive des ressources humaines en utilisant les algorithmes du Machine Learning dans l'analyse de turn-over et la prédiction des employés ayant tendance de quitter leurs emplois en se basant sur une base de données de l'entreprise IBM publiée dans le cadre d'une compétition pour le développement d'un modèle interne qui sert à identifier les employés ayant l'intention de quitter leurs emplois. Les résultats de cette étude ont montré que cette technologie peut jouer un rôle crucial dans la gestion proactive des ressources humaines en offrant des informations qui permettent de proagir et anticiper les actions liées à la gestion des ressources humaines.

**Mots clés :** Machine Learning, La gestion proactive des ressources humaines, Turn-over.

**Classification JEL :** M12

**Type de l'article :** Recherche appliquée

### Abstract:

Since the Industrial Revolution the function of human resources (HR) has undergone several changes, today it is asked to adopt a long-term and proactive approach which serves to anticipate the needs of the company and the problems likely to impact its productivity and performance in order to implement long-term adaptation actions and be effective in its strategic approach. Otherwise Machine learning has been for some time, a trending technology that has seen massive use in many areas, according to a study conducted by IT decision makers from over 15 different business sectors in the UK, France, Germany and Spain, 87% of the samples have implemented this technology or plan to do so. This technology has allowed companies to improve their processes, increase their competitiveness and help decision-making in several management areas such as finance and marketing.

This present work seeks to highlight the potential of this technology to promote proactive management of human resources by using Machine Learning algorithms in the analysis of turnover and the prediction of employees tending to leave their jobs using a IBM corporate database published as part of a competition for the development of an internal model used to identify employees intending to leave their jobs. The results of this study have shown that this technology can play a crucial role in the proactive management of human resources by providing information that makes it possible to pro-act and anticipate actions related to human resources management.

**Key-word:** Machine learning, proactive management of human resources, Turn-Over

**JEL classification:** M12

**Paper type:** Empirical research

## 1. Introduction.

Dans le contexte actuel, marqué par l'évolution technologique continue, l'instabilité de l'environnement et l'intensification de la concurrence. La fonction de ressources humaines (RH) est sollicitée à jouer un rôle proactif permettant à l'entreprise d'achever leurs objectifs stratégiques et acquérir un avantage compétitif vis-à-vis des concurrents. Le recours aux Machines Learning (ML) comme un levier de gestion proactive dans plusieurs domaines s'avère être une solution innovante pour améliorer les processus, augmenter la productivité et la compétitivité et aider à la prise des décisions, ce qui explique l'investissement des entreprises jusqu'à 37,5 milliards de dollars dans des logiciels d'apprentissage automatique en 2019, et jusqu'à 97,9 milliards de dollars en 2023 selon les prévisions de l'international data corporation (IDC).

La valeur ajoutée des techniques de ML dans plusieurs domaines de gestion est prouvée par plusieurs travaux tels que la finance où les algorithmes de ML sont utilisés dans la gestion de portefeuille, l'identification des risques de crédit et l'allocation d'actifs (Caiazza & al.,2020). De même le ML dernièrement a connu une utilisation massive dans le marketing notamment, au niveau de la prédiction de comportement des consommateurs, l'analyse de la structure du marché, le développement des systèmes de recommandation, etc.(Arasu & al.,2020, Ma & Sun, 2020, Breni,2020). Tandis que l'utilisation de ces techniques dans la gestion des ressources humaines a fait l'objet d'un nombre réduit d'études, pour cela ce présent article a pour objet de démontrer que **comment les algorithmes du ML peuvent favoriser la gestion proactive des ressources humaines ? Et dans quelle mesure ces techniques peuvent contribuer à l'amélioration de la gestion des ressources humaines ?**

Pour répondre à ces questions nous avons dans un premier lieu mobilisé deux revues de littérature qui traitent les concepts de base concernant la Machine Learning et l'approche proactive de gestion des ressources humaines, puis nous avons essayé de mettre en évidence la relation entre les deux concepts à travers une étude de cas dans laquelle nous avons développé une Machine Learning permettant d'analyser le turn-over et prédire les employés ayant tendance de quitter leurs emplois en se basant sur une base de données interne de l'entreprise IBM.

## 2. La gestion des ressources humaines vers une approche proactive.

La fonction Rh a connu plusieurs changements depuis la révolution industrielle grâce à l'évolution de la perception de l'homme dans l'entreprise d'un moyen à gérer à une ressource à développer, son rôle dans la création de la valeur vu l'accroissement de travail intellectuel par rapport au travail physique et à la dynamique de l'environnement. À cet égard on peut distinguer entre trois transformations majeures de la fonction Rh à savoir ; la gestion de personnel, la gestion des ressources humaines et la gestion stratégique des ressources humaines (Chaminade B, 2008).

En effet depuis la révolution industrielle, la fonction production est considérée comme la seule fonction créatrice de valeur et la gestion du personnel n'était qu'une fonction d'exécution des objectifs de l'organisation de travail pour faire face aux besoins de la production ce qui explique l'émergence des méthodes et procédures de travail qui standardise l'activité humaine en fonction des méthodes de production, notamment l'organisation scientifique de travail (OST). Un peu plus tard le concept de la gestion des ressources humaines a vu le jour suite à l'essor de l'école des relations humaines qui a mis en évidence le rôle de l'humain dans la création de la valeur au sein de l'entreprise ainsi pour faire face aux grèves engendrées par les conditions de travail inappropriées. Par ailleurs le passage vers une gestion stratégique des ressources humaines et l'intégration de la fonction RH dans le processus stratégique des entreprises est dû à la nécessité pour celles-ci d'avoir une vision globale des ressources humaines et de les intégrer

aux principaux enjeux organisationnels (G.Gurin,2002). Cette intégration est justifiée aussi par l'environnement turbulent dans lequel évoluent les organisations aujourd'hui et à la contribution des ressources humaines au succès (ou à l'échec) des stratégies et de développement de ces organisations (Guerin & wils,2002).

Pour cela le rôle actuel de la fonction RH a pris de plus en plus une dimension stratégique et qui consiste à fournir un capital humain en termes de qualité et quantité permettant de répondre aux enjeux stratégiques de l'entreprise et achever leurs objectifs à long terme, ainsi prévenir les problèmes d'ordre Rh et mettre en place des plans d'actions pertinentes pour y faire face afin d'accroître la performance du capital humain et contribuer positivement à la compétitivité de l'entreprise(Peretti, 2019).

Ce nouveau rôle souligne l'importance de l'approche proactive dans la gestion des ressources humaines qui s'articule autour de l'anticipation des besoins futurs en compétences, ce qui permet de proagir sur plusieurs volets Rh entre autres, la gestion des compétences, la formation, le recrutement, le système de gestion de la relève, etc. Ainsi comprendre et anticiper son environnement, notamment les facteurs qui impacte la productivité des employés et les problématiques Rh telles que le turn-over, l'absentéisme, l'engagement organisationnel, etc.

### **3. Machine Learning.**

L'autoapprentissage « Machine Learning » est un domaine fascinant issu de différentes disciplines telles que les statistiques, l'optimisation, l'algorithmique ou le traitement du signal. C'est un champ d'études en développement continu et qui prend aujourd'hui une position importante dans l'ensemble des domaines. En effet le Machine Learning a connu une utilisation massive depuis quelques années dans la reconnaissance automatique de caractères ou les filtres antispam, il sert maintenant à protéger contre les fraudes bancaires, la recommandation des produits, vidéos, films ou autres produits adaptés à nos goûts, la prédiction des ventes, l'analyse des cours boursiers, etc. Son impact est aussi fortement notable dans d'autres domaines, tels que les voitures autonomes et dans l'informatique utilisée dans les bâtiments ( Boutaba & al., 2018)

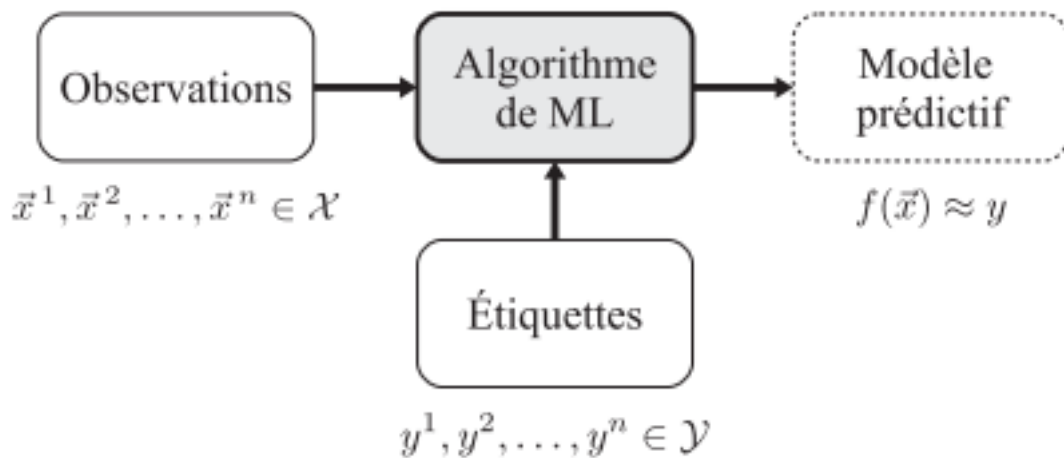
L'apprentissage au sens large est une modification d'un comportement sur la base d'une expérience (Benureau, 2015). Dans le cas d'un programme informatique, signifie la capacité d'un programme d'apprendre sans que cette modification ne soit explicitement programmée (Arthur, 1959). En effet le Machine Learning est une sous-branche de l'intelligence artificielle désignant l'ensemble des méthodes et algorithmes permettant à des machines de découvrir et d'apprendre des modèles, sans que ces dernières ne reçoivent d'instruction de programmation ( Rasekhschaff & Jones, 2019), notamment tirer les relations et les liens entre des variables qui font l'objet d'une large base de données, où plus de données sont nombreuses plus le Machine Learning peut être exploitée efficacement ( Le bigdata, 2018).

En effet le Machine Learning permet de répondre à plusieurs problématiques de différents types en mobilisant un ensemble d'algorithmes qui peuvent être regroupés sous forme de différentes class, entre autre l'apprentissage supervisé et l'apprentissage non supervisé.

#### **3.1 L'apprentissage supervisé**

Dans le cadre de l'apprentissage supervisé, on essaie de définir une règle de prédiction d'une variable à prédire Y (étiquettes) en fonction de variables prédictives X (observations), on dispose pour cela des données pour lesquelles à la fois X et Y sont observés et on cherche parmi une famille de règles possibles, celle qui optimise un critère de qualité à définir. L'objectif est ensuite de pouvoir appliquer le modèle prédictif à de nouvelles données pour lesquelles seuls X est connu afin d'en déduire une prédiction Ypred (Birmelé,2020).

Figure 1: Apprentissage Supervisé



Source: Azencott, 2018

Ce type d'apprentissage fait l'objet de plusieurs algorithmes qui permettent de développer un modèle prédictif à travers l'analyse des relations existant entre des observations et des étiquettes, entre autre ;

### 3.1.1 La régression linéaire.

La régression linéaire (simple ou multiple) est l'un des algorithmes d'apprentissage supervisé les plus populaires, qui sert à expliquer une variable Y à l'aide d'une ou plusieurs variables X. la variable y est appelée variable dépendante ou variable à expliquer et les variables x sont appelées variables indépendantes ou variables explicatives (Cornillon & Matzner, 2010).

### 3.1.2 La régression logistique.

La régression logistique est un algorithme de classification qui permet d'expliquer une variable d'intérêt binaire (Oui / Non, Vrai / Faux, Spam / Ham), tandis que les variables explicatives peuvent prendre une forme quantitative ou qualitative (Lutz & al., 2015).

La classification par régression logistique vise à déterminer la probabilité que la variable de sortie appartienne à une classe spécifique en fonction d'une sommation linéaire des caractéristiques. Cette fonction est affirmée en (1) et (2).

$$z = w_0 + w_1X_1 + w_2X_2 + \dots + w_pX_p \quad (1)$$

$$P(G = 1) = \frac{e^z}{e^z + 1} \quad (2)$$

Sachant que  $X_1 \dots X_p$  correspond aux variables, omégas représentent les coefficient et  $P(G=1)$  est la probabilité que G appartient à la classe 1(Friedman et al., 2001).

### 3.1.3 K-NN

Le classificateur K-NN ( K- Nearest Neighbors) consiste à classer les observations non étiquetées en les affectant à la classe des exemples étiquetés les plus similaires, les caractéristiques des observations sont collectées à la fois pour l'ensemble de données d'apprentissage et de test. La fonction K-NN peut être déterminée selon l'équation suivante (Zhang, 2016) ;

$$D(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

### 3.2 L'apprentissage non supervisé.

L'apprentissage non supervisé consiste à découvrir des structures cachées dans un ensemble de données sans étiquettes, dans ce cas le résultat souhaité est inconnu. Ce type d'apprentissage inclus un ensemble d'algorithmes tels que, Apriori, Clustering. (Fleischer, 2019).

#### 2.2.1 Clustering.

Le Clustering a pour objectif de distinguer entre des groupes d'observations homogènes, de sorte que les observations au sein d'un groupe soient les plus semblables possibles et les groupes soient les plus différents possibles les uns des autres.

Cet algorithme est généralement basé sur une notion de la distance euclidienne sur des données normalisées :

$$d(X_i, X_j) = \left( \sum_{\ell=1}^p (X_{i\ell} - X_{j\ell})^2 \right)^{\frac{1}{2}}$$

#### 2.2.2 Apriori

Apriori est le premier algorithme de recherche de règles d'association incluant des étapes d'élagage pour tenir compte de la croissance exponentielle du nombre d'ensembles d'items candidats (Argawal & Srikant, 1994). Cet algorithme est couramment utilisé dans l'analyse des paniers de marchés, où l'on vérifie les combinaisons de produits qui apparaissent fréquemment dans la base de données.

Notons  $C_k$  l'ensemble des ensembles d'items candidats de la taille  $K$  et  $F_k$  l'ensemble d'items fréquents de taille  $k$ .

- L'algorithme commence par déterminer le support de chaque item : initialisation de  $F_1$ .
- L'algorithme génère itérativement les ensembles d'items fréquents de taille  $k$  à partir des ensembles d'items de taille  $(k - 1)$  obtenus à l'étape précédente.

## 4. Méthodologie

### 4.1 Présentation de cas

Afin de démontrer l'impact de la machine Learning sur l'amélioration de la gestion proactive des ressources humaines, dans ce qui suit, nous allons essayer de développer une Machine Learning permettant de comprendre et anticiper des problématiques liées à la gestion des ressources humaines. Étant donné que la base de données reste un élément indispensable pour entraîner et tester notre machine Learning et vu que les entreprises n'ont pas tendance d'offrir les données nécessaires pour une telle étude pour motif de confidentialité, nous allons utiliser une base donnée publiée par l'entreprise IBM téléchargée d'auprès le site <https://www.kaggle.com/>, et qui sert à analyser le turn-over et prédire les employés ayant tendance de quitter leurs emplois.

Alors notre objectif sera le développement des algorithmes permettant d'identifier les variables ayant un effet sur la décision des employés d'IBM de rester ou quitter et aussi prédire les employés qui vont démissionner.

### 4.2 Méthodologie de traitement de données.

Le turnover reste parmi les problèmes majeurs pour la gestion des ressources humaines qui entraîne un coût financier important pour l'entreprise (perte de productivité du salarié, perte des investissements en termes de formation, le coût lié à l'embauche et l'intégration d'un nouveau salarié, etc.). De même le départ d'un salarié est rarement anticipé, ce qui est susceptible à engendrer une perturbation de l'organisation de l'entreprise.

En outre, le turnover représente une perte de compétences pouvant impacter la qualité de travail selon son importance et la rareté de profile, ainsi il peut même nuire à l'image de l'entreprise et faire perdre des clients. Pour cela l'anticipation des employés ayant tendance de quitter l'entreprise et identifier ce qui favorise leur démission aidera les responsables RH à mettre en place des actions efficaces pour retenir leurs employés ou bien préparer leurs départs à l'avance pour minimiser les effets de départ non planifié et même orienter le recrutement et la formation en privilégiant les candidats qui vont rester la période la plus longue au sein de l'entreprise. L'analyse de cette base de données aura l'objet d'explorer la relation entre les caractéristiques des employés et leurs décisions de rester ou quitter, l'identification des caractéristiques ayant un impact important sur cette décision et finalement le développement d'un modèle qui nous permettra de prédire la décision des autres employés.

Pour ce faire notre analyse sera principalement basée sur un algorithme d'apprentissage supervisé, à savoir, la régression logistique en utilisant l'outil R Studio.

Pour l'aboutissement de notre analyse, nous allons poursuivre les étapes suivantes ;



### 4.3 Aperçu sur la base de données.

IBM Attrition Data Set est une base de données interne créée par l'entreprise IBM pour développer un modèle prédictif qui permet d'anticiper le turnover chez eux. La base de données contient des informations concernant 1471 employés en prenant en considération 34 variables liées aux informations personnelles basiques et l'expérience au travail des employés qui ont déjà quitté et qui sont encore en poste (Annexe 1).

Tableau 1: Aperçu sur la base de données

Variable	Description
Age	Valeur numérique
TurnOver	Employé quitté (0=Non, 1=Oui)
Voyage professionnel	(1=Pas du tout, 2=Fréquemment, 3=Rarement)
Salaire horaire	Valeur Numérique
Département	(1=RH, 2=R&D, 3=Commerciale)
Distance de son foyer	Valeur Numérique- La distance entre le foyer et le travail
Éducation	Valeur Numérique
Domaine d'étude	(1=RH, 2= Sciences De Vie, 3=Marketing, 4=Sciences médicales, 5=autres, 6= Technique)

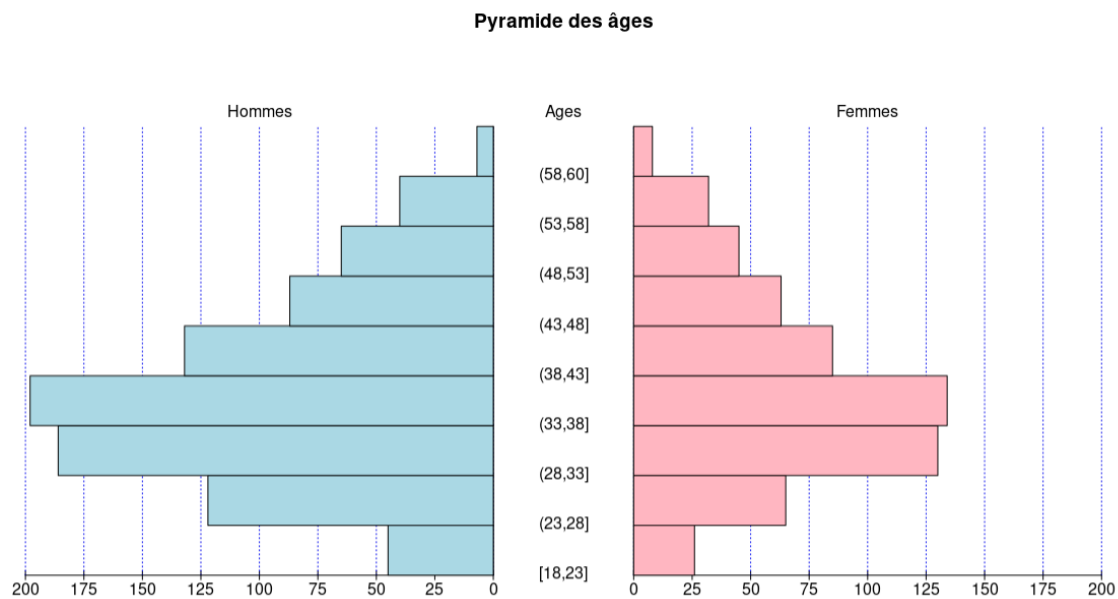
Source : Les auteurs

### 4.4 Aperçu générale sur les indicateurs RH.

Dans cette partie nous allons essayer de visualiser les informations générales sur notre population, telles que la répartition socio-professionnelle, le niveau de formation et de rémunération, la satisfaction, etc.



Figure 2: La pyramide des âges



Source: Les auteurs

On remarque que la tranche d'âge la plus représentée dans notre population est celle entre 28 et 38 ans, ce qui donne la forme en « poire écrasée » qui correspond classiquement à une entreprise ayant cessé de recruter durant de longues années (figure 2). Dans ce cas le risque de pertes de savoirs et d'expériences est très important, surtout dans le secteur de hautes technologies, qui peuvent en résulter des difficultés de renouvellement des connaissances en matière technologique.(Godelier,2007).

Figure 2: La densité des revenus mensuels en fonction du genre

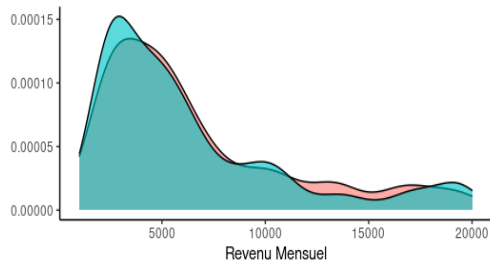


Figure 3: Effectif par Niveau d'éducation

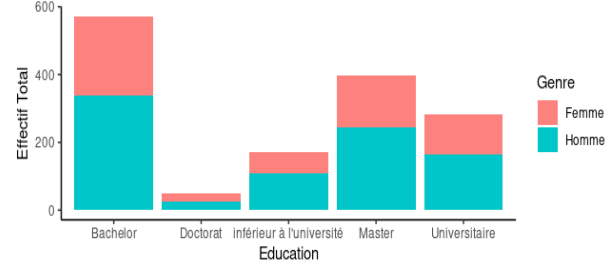
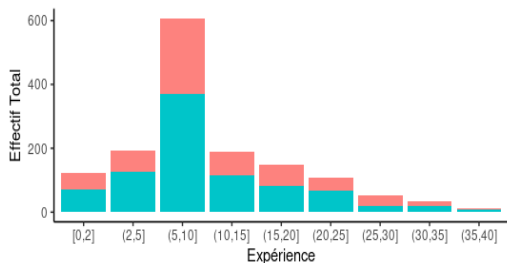
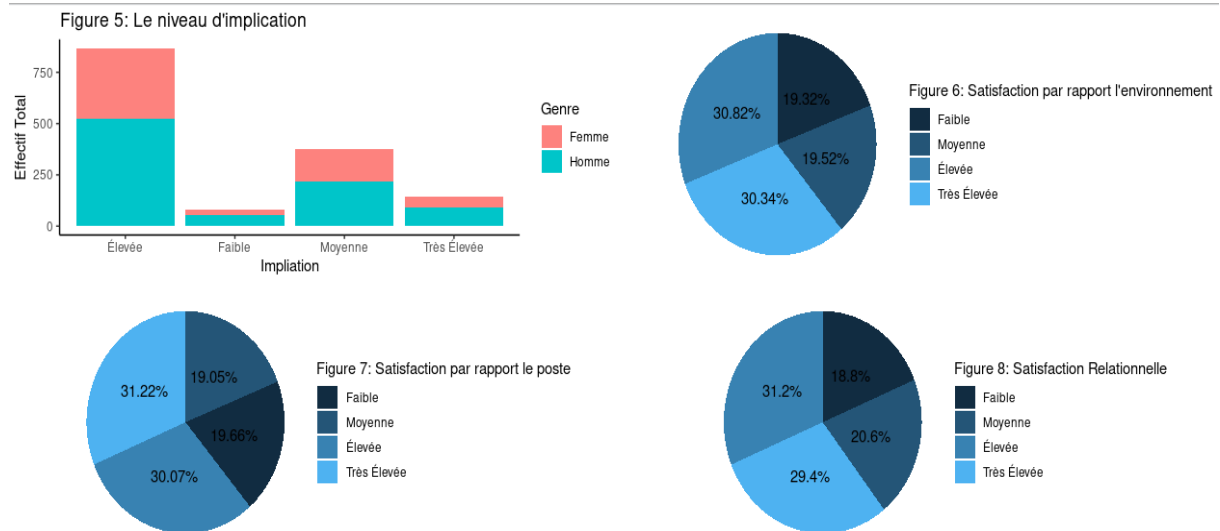


Figure 4: Effectif par Niveau d'expérience



Source : Les auteurs

Selon la figure 2 on constate qu'une grande partie des employés touchent un salaire entre 1000 et 5000 dollars, ce qui semble un peu logique vu que la plupart ont une expérience inférieure ou égale à 10 ans (figure 3), ainsi plus que le moitié des employés a un niveau d'éducation inférieur ou équivalent au Bachelor (figure 4).



Source : Les auteurs

L'implication des employés signifie le degré d'attachement d'un individu à son organisation, et qui impact positivement la décision de l'individu de rester ou quitter son emploi (Gutiérrez-Martinez, 2006). Pour les employés d'IBM on observe que 59.05% des employés ont un niveau d'implication élevé, tandis que 5.64%, 25.51%, 9.80 % des employés ont un niveau d'implication faible, moyen et très élevé (figure 5). Par ailleurs la satisfaction au travail reste parmi les facteurs qui favorisent la rétention des employés (Meyer& al.,1993), cependant on observe que 19.32 % ont une satisfaction faible par rapport à l'environnement (Figure 6), 18.8% ont le même niveau de satisfaction vis-à-vis les relations intra-personnelles (Figure 7), et 19,05% ont une faible satisfaction par rapport le poste (Figure 8).

## 5. Développement de modèle et résultats

### 5.1 La restructuration des données

Avant de développer notre modèle prédictif en se basant sur l'algorithme de régression logistique, il semble important de convertir les variables catégorielles en variables numériques afin d'être prises en considération par notre algorithme. À titre d'exemple la variable « la situation familiale » sera transformée en trois variables binaires suivantes ; Situation\_familliale\_Marié (0 , 1), Situation\_familliale \_Divorcé(0,1) et Situation\_familliale \_célibataire (0,1) où 1 et 0 signifier simultanément Vrai et Faux. (Tableau 2).

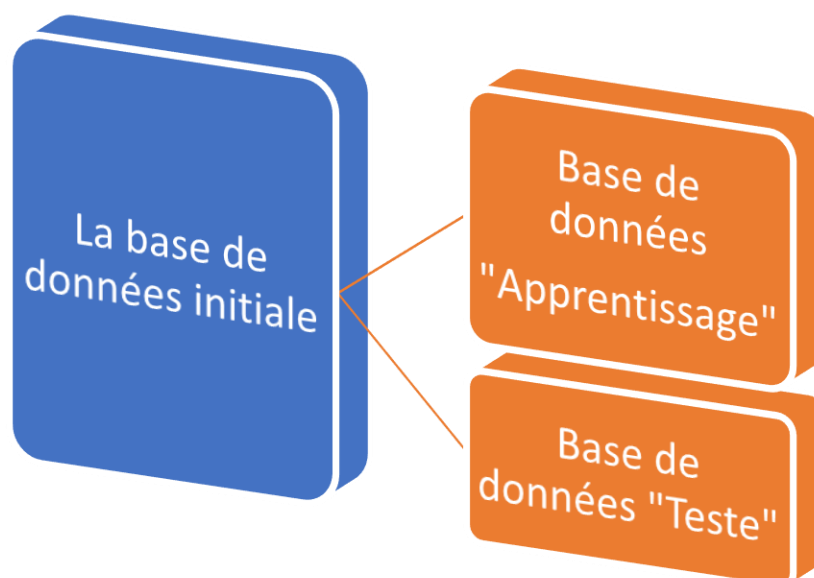
Tableau 2: Exemple de la transformation des variables catégorielles en variables binaires.

	Situation_familliale_Marié	Situation_familliale _Divorcé	Situation_familliale _célibataire
Marié	1	0	0
Divorcé	0	1	0
Célibataire	0	0	1

Source : Les auteurs

Après avoir assuré l'adéquation du type des variables aux Machines Learning, nous avons divisé notre base de données en deux bases de données sous-jacentes « Apprentissage » que nous allons utiliser pour le développement de notre modèle et « Teste » sur laquelle nous allons tester la performance de notre modèle et sa capacité prédictive, et qui continent simultanément 75% et 25% de la base de données initiale.

Figure 9: La division de la base de données



Source : Les auteurs

## 5.2 La sélection des variables.

Afin d'avoir un modèle prédictif efficace, il est important d'éviter les variables qui n'ont aucun effet sur la variable étudiée. Pour cela nous avons écarté un ensemble de variables tels que, « Supérieur à 18 » vu que l'ensemble des employés sont supérieurs à 18 ans, « ID » qui n'aura aucun effet sur leurs décisions de quitter ou de rester dans l'entreprise et « Horaire standard » parce que l'ensemble des employées ont un nombre d'heures standards similaire.

Par ailleurs l'obtention d'un modèle pertinent nécessite l'élimination des colinéarités notamment les variables dépendantes ayant une forte corrélation linéaire entre eux (C.F.Dormann & al.), ainsi celles qui ne représentent aucun effet significatif sur la variable indépendante. Pour cela nous avons établi la procédure Forward Stepwise Selection<sup>1</sup> pour recenser les variables qui permettent d'améliorer notre modèle en se basant sur le critère AIC<sup>2</sup>.

## 5.3 Le modèle final.

Selon les résultats de la procédure Forward Stepwise Selection nous avons abouti au modèle ci-dessous (Tableau 3) composé par 31 variables indépendantes avec un niveau d'importance disparate et un type de relation différente. Comme montré dans le tableau ci-dessous, on constate que les variables surlignées en rouge correspondent aux variables qui favorisent le turnover au sein d'IBM notamment, les heures supplémentaires, les voyages professionnels d'une cadence fréquente, le nombre d'entreprises dans lesquelles l'employé a déjà travaillé plus qu'il est important plus que la probabilité qu'il quitte IBM augmente, le nombre d'années depuis la dernière promotion, l'ancienneté et ainsi la distance entre le foyer et le travail. Tandis que les variables qui contribuent à la rétention des employés sont nombreuses entre autres, le travail en tant que directeur de recherche, représentant médical, directeur de production, l'implication, l'équilibre vie privée, la satisfaction au travail, le niveau des stock-options, etc.

<sup>1</sup> C'est un type de régression pas à pas qui commence avec un modèle vide et ajoute les variables une par une. À chaque étape, vous ajoutez la seule variable qui donne la meilleure amélioration à votre modèle.

<sup>2</sup> Le critère d'information d'Akaike (AIC) est une mesure qui permet de comparer entre les modèles statistiques, calculé à partir du nombre de paramètres à estimer du modèle et du maximum de la fonction de vraisemblance.

**Tableau 3: Le modèle final de régression logistique**

<i>Variabes</i>	<i>Coeffi- cient</i>	<i>Std.Error</i>	<i>z value</i>	<i>Pr(&gt; z )</i>	
<i>Constante</i>	5,69321	0.88364	6.443	1.17e-10	***
<i>Heures supplémentaires</i>	1,83740	0.21782	8.435	< 2e-16	***
<i>Voyage professionnel fréquent</i>	0,58076	0.24439	2.376	0.017484	*
<i>Nombre d'anciens employeurs</i>	0,19826	0.04254	4.660	3.15e-06	***
<i>Nombre d'années depuis la dernière promotion.</i>	0,14674	0.04650	3.155	0.001602	**
<i>Ancienneté au sein de l'entreprise</i>	0,09336	0.04281	2.181	0.029207	*
<i>La distance de son foyer</i>	0,03448	0.01225	2.816	0.004867	**
<i>Poste_Directeur de recherche</i>	-2,15461	0.85279	- 2.527	0.011519	*
<i>Poste_Représentant médical</i>	-1,76649	0.50409	- 3.504	0.000458	***
<i>Poste_Directeur de production</i>	-1,51557	0.45040	- 3.365	0.000766	***
<i>Situation familiale _Divorcé</i>	-1,36474	0.38652	- 3.531	0.000414	***
<i>Poste_Manager</i>	-1,27546	0.66660	- 1.913	0.055700	.
<i>Poste_Chercheur</i>	-1,05350	0.28066	- 3.754	0.000174	***
<i>Voyage professionnel du tout</i>	-0,95075	0.40994	- 2.319	0.020382	*
<i>Situation familiale _Marié</i>	-0,84682	0.27426	- 3.088	0.002018	**
<i>Poste_Directeur des ventes</i>	-0,62012	0.27391	- 2.264	0.023577	*
<i>Formation initiale_Sciences de la vie</i>	-0,57397	0.24701	- 2.324	0.020144	*
<i>Formation initiale_médical</i>	-0,51284	0.26165	- 1.960	0.049996	*
<i>Implication</i>	-0,47107	0.13573	- 3.471	0.000519	***
<i>Équilibre vie pro-privé</i>	-0,43990	0.14128	- 3.114	0.001847	**
<i>Satisfaction par rapport au poste</i>	-0,43885	0.09249	- 4.745	2.09e-06	***

<i>Satisfaction par rapport à l'environnement</i>	-0,39914	0.09292	-	1.74e-05	***
			4.295		
<i>Satisfaction par rapport au relations</i>	-0,24303	0.09360	-	0.009418	**
			2.596		
<i>Niveau StockOption</i>	-0,18436	0.17307	-	0.286774	
			1.065		
<i>Nombre d'heures de formation la dernière année</i>	-0,16330	0.08207	-	0.046608	*
			1.990		
<i>Années avec le manager actuel</i>	-0,12049	0.05200	-	0.020491	*
			2.317		
<i>Années dans le poste actuel</i>	-0,10764	0.05050	-	0.033042	*
			2.132		
<i>Ancienneté totale</i>	-0,08639	0.03048	-	0.004598	**
			2.834		
<i>Age</i>	-0,01826	0.01452	-	0.208469	
			1.258		

Source : Les auteurs.

#### 5.4 L'évaluation de modèle

Pour tester la capacité prédictive de ce modèle, nous avons essayé de prédire les employés qui vont quitter leurs emplois en utilisant la base de données tests que nous n'avons pas utilisée pendant la phase de l'apprentissage supervisé. L'application de la Machine Learning sur un échantillon de 367 employés nous a donné la matrice de confusion ci-dessous.

Tableau 4: La matrice de confusion entre la prédiction de l'algorithme et les résultats réels.

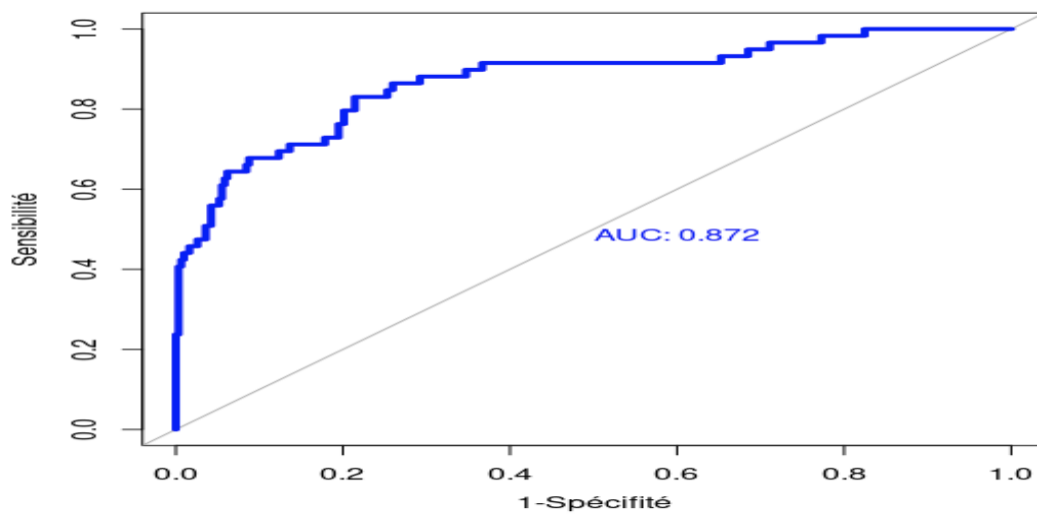
	Turn-Over	La prédiction de l'algorithme	
		Non	Oui
La réalité	Non	242	66
	Oui	10	49

Source: Les auteurs

Les résultats obtenus montre que notre Machine Learning a pu prédire jusqu'à 49 employés quittés parmi 59 (83%) avec un taux de précision de 42% ( $49/(49+66)$ ) et un espace sous la courbe égale à 87,2% (AUC)<sup>3</sup> (Figure 10).

<sup>3</sup> AUC: il s'agit de la zone sous la courbe de caractéristique de fonctionnement du récepteur (ROC), qui permet de mesurer la performance du modèle à distinguer entre les classes, qui représentent dans notre cas la décision de l'employé de quitter ou de rester. Sachant que le modèle parfait correspond à un AUC=100%

Figure 10 : le courbe ROC

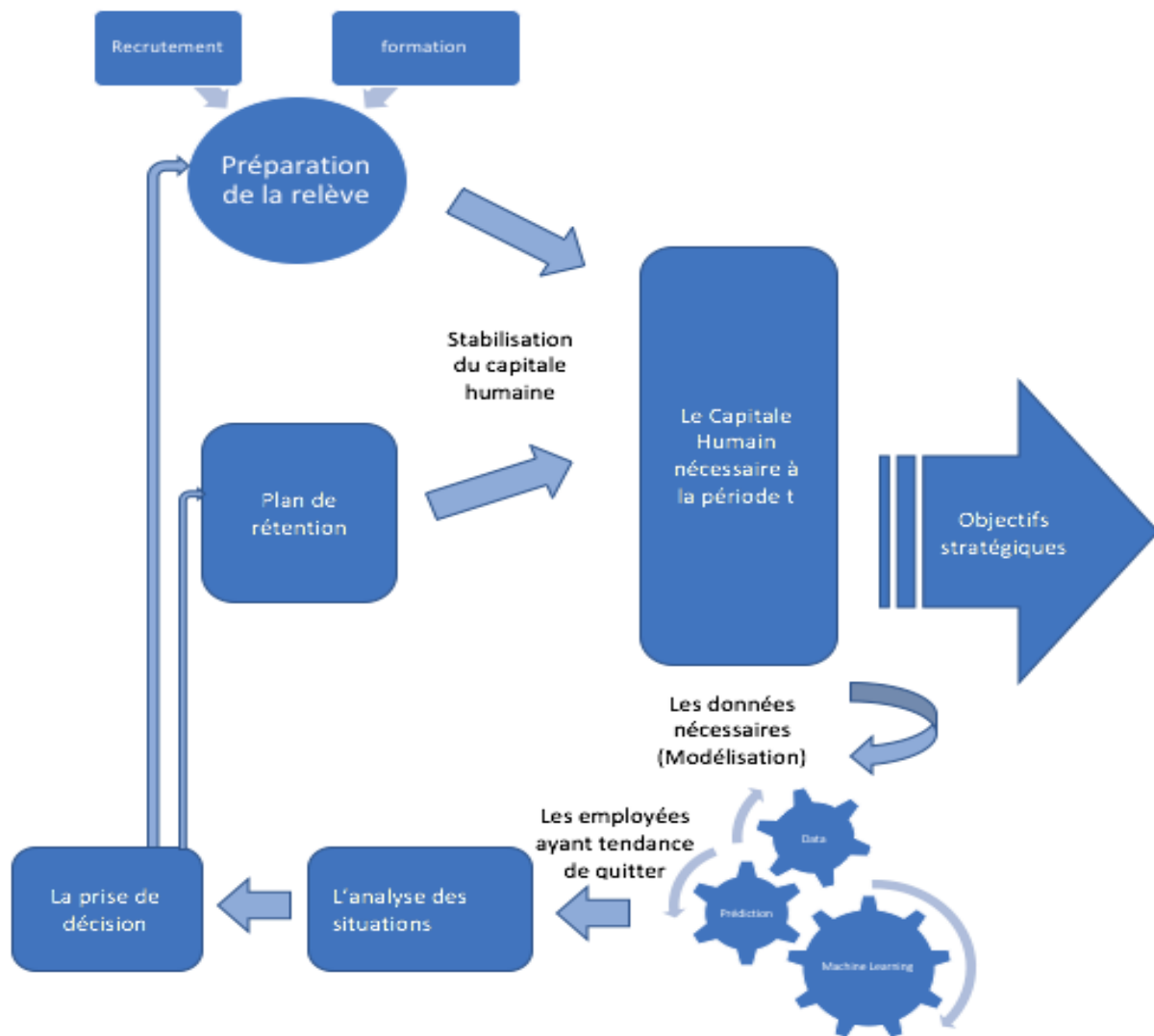


Source : les auteurs

## 6. Discussion

À partir de ce modèle obtenu par l'un des algorithmes d'apprentissage supervisé, à savoir la régression logistique, nous avons essayé de mettre en évidence le potentiel de cette technologie pour favoriser une gestion proactive des ressources humaines, en vue de contribuer à la réalisation des objectifs stratégiques de l'entreprise et avoir un avantage compétitif par rapport aux concurrents. Les départs non planifiés sont capables de perturber le capital humain de l'entreprise en termes de quantité et qualité et ralentir son évolution, de même ceci génère une mauvaise image de marque employeur, ce qui impact négativement l'attractivité des talents par rapport aux concurrents. Le recours aux Machines Learning pour prédire les employés ayant tendance de quitter leurs emplois peut être considéré comme une solution magique permettant à l'entreprise de proagir et mettre en place les actions adéquates pour chaque employé ayant tendance de quitter, ce qui est confirmé par notre étude de cas où nous sommes arrivés à détecter 83 % des employés quittés. Les informations d'écoulées de notre modèle prédictif permet à l'entreprise de prendre des décisions pertinentes en fonction de chaque employé ayant tendance de quitter son emploi, que ce soit un plan de rétention lorsqu'il s'agit d'un talent ayant une forte valeur ajoutée, ou laisser faire au cas contraire, ainsi préparer à l'avance le départ de l'employeur surtout lorsqu'il s'agit d'un poste critique qui consiste à chercher un profil très rare ou même former la relève en interne, alors on remarque que le départ d'un employé engendre une mobilisation de l'ensemble des processus RH et nécessite du temps pour stabiliser le capital humain dont l'entreprise a besoin, par ailleurs cette information fournie par le Machine Learning permette de réduire le temps dédié à cette stabilisation et donc répondre au besoin stratégique de l'entreprise en matière des ressources humaines.

Figure 11 : Schémas synthétiques de la gestion proactive du Turn-over basé sur le Machine Learning.



Source : Les auteurs

## 7. Conclusion

L'objectif de cette étude est de mettre en évidence l'utilité des Machines Learning au sens étroit et l'intelligence artificielle au sens large dans la gestion proactive des ressources humaines, à cette fin nous avons développé un modèle prédictif en nous basant sur l'un des algorithmes de la Machine Learning pour comprendre et prévenir le Turn-Over qui est considéré comme l'un des problèmes les plus importants au niveau de la gestion des ressources humaines. Cette machine Learning nous a donné des résultats prometteurs où notre algorithme arrive à prédire jusqu'à 83% des employés quittés avec un taux de précision égale à 42%. Ce qui va permettre aux responsables des ressources humaines d'anticiper ce départ en essayant de retenir l'employé, en agissant sur les facteurs les plus importants qui impactent son départ ou bien préparer à l'avance la relève qui va occuper le poste surtout lorsqu'il s'agit d'un poste critique. Ainsi la prédiction faite par notre Machine Learning à ce propos peut contribuer à la prise de décision sur de nombreux volets RH tels que, la formation, la GPEC, la gestion des carrières, la sélection des candidats, etc.

Par rapport au cas du turn-over, il semble que la Machine Learning permet de favoriser une gestion proactive des ressources humaines en offrant des informations avec une précision acceptable qui permettent d'anticiper plusieurs problèmes en agissant tôt, ce qui souligne

l'importance de l'utilisation de ces techniques pour améliorer la gestion des ressources humaines et contribuer à la performance et la compétitivité de l'entreprise. Par ailleurs il semble important d'explorer les problématiques Rh qu'on peut traiter par cette technique afin d'enrichir le débat sur la capacité de ces algorithmes à favoriser la proactivité en termes de la gestion des ressources humaines.

## Références

- (1) Agrawal, R., et Srikant, R. (1994, September). *Fast algorithms for mining association rules*. In *Proc. 20th int. conf. very large data bases*. (1215). 487-499.
- (2) Allen, N. J., Meyer, J. P., (1990b). *Organizational socialization tactics : A longitudinal analysis of links to newcomers' commitment and role orientation*. *Academy of Management Journal*. 4(33). 847-858.
- (3) Arasu, B., Seelan, J., Thamaraiselvan, N., (2020). *A machine learning- based approach to enhancing social media marketing* . *Computers and electronical engineering*. (86). Article 106723.
- (4) Azencott, C.A., (2018). *Introduction au machine learning*, Paris : Dunod.
- (5) Benureau, F., (2015),. *L'apprentissage est une modification d'un comportement sur la base d'une expérience*. Paris : Dunod.
- (6) Birmelé, E. (2020). *Apprentissage supervisé*. <https://helios2.mi.parisdescartes.fr/~ebirmele/depots/Enseignements/Strasbourg/apprentissage.pdf>
- (7) Boutaba, R., Salahuddin, M. A., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F., & Caicedo, O. M. (2018). *A comprehensive survey on machine learning for networking: evolution, applications and research opportunities*. *Journal of Internet Services and Applications*, 9(1), 1-99.
- (8) Brei, V., (2020), *Machine learning in marketing: Over view, learning strategies, Applications, and Future developpements*. *Foundations and Trends in marketing*. 3(14). 173-236.
- (9) Caiazza, A., Champassak, C., Floch, G., Salem, H. (2020). *Le machine Learning en finance :Vers une véritable révolution*. Paris :Square.
- (10) Chaminade, B., (2008). *Guide pratique Rh et compétences*. Paris : Edition Afnor
- (11) Cornillon, P.A, et Matzner, E., (2010). *Régression avec R*, Berlin : Springer.
- (12) F.Dormann, C., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al., (2013). *Collinearity: a review of methods to deal with it and a simulation study evaluating their performance*. *Ecography*. 1(36). 27-46.
- (13) Fleischer, D. (2019). *Automated Trading System using Machine Learning*. thèse de doctorat. Univeristé de Stavanger. Norvège.
- (14) Friedman, J., Hastie, T., Tibshirani, R., (2001). *The Elements of Statistical Learning*; Series in Statistics. New York: Springer.
- (15) Godelier, E., (2007). *Pyramide des âges et gestion des ressources humaines*. Vingtième siècle. *Revue d'histoire*. 3(95). 127-142.
- (16) Guerin G.,(2002). *Les tendances en gestion stratégique des ressources humaines*. Avis d'expert dans Dolan, Saba, Jakson et Schuler, la gestion des ressources humaines : tendances, enjeux et pratiques actuelles. 3<sup>e</sup> édition
- (17) Guerin, G., et Wils, T., (2002). *la gestion stratégique des ressources humaines*, *Gestion*. 2(27).14-23
- (18) Gutiérrez-Martinez, I., (2006). L'implication organisationnelle des professionnels de TI. *Revue française de gestion*. 9(168). 143-156.



- (19) Lebigdata. (2018). *Machine Learning et Big Data : définition et explications*. Consulté sur <https://www.lebigdata.fr/machine-learning-et-big-data>
- (20) Ma, L., Sun, B., (2020). *Machine learning and AI in marketing- Connecting computing power to human insights*. International Journal of research in marketing. 3(37). 481-504.
- (21) Peretti, J-M., (2016). *La gestion des ressources humaines*. Paris :Vuibert. (Édition 21)
- (22) Rasekhschaffe, K. C., et Jones, R. C. (2019). *Machine learning for stock selection*. Financial Analysts Journal, 75(3), 70-88.
- (23) Samuel, A. L., (1959). *Some studies in machine learning using the game of checkers*. IBM Journal of Research and Development, 3(3), 211–229.
- (24) Zhang, Z. (2016). *Introduction to machine learning: k-nearest neighbors*. Annals of translational medicine, 4(11).

**Annexe 1 : Les composants de la base de données IBM**

Nom	Description
Age	Valeur numérique
TurnOver	Employé quitte l'entreprise (0=non, 1=Oui)
Voyage professionnel	(1=Pas du tout, 2=Fréquemment, 3=Rarement)
Salaire Journalier	Valeur Numérique - Le niveau de salaire
Département	(1=RH, 2=R&D, 3=Commerciales)
Distance de son foyer	Valeur Numérique- La distance entre le foyer et le travail
Éducation	Valeur numérique- 1=Inférieure à l'universitaire ,2= Universitaire ,3=Bachelor ,4=Master ,5=Doctorat
Domaine d'étude	(1=RH, 2= Sciences De Vie, 3=Marketing, 4=Sciences médicales, 5=autres, 6= Technique)
ID	Valeur numérique - Identification
Satisfaction Environnement	Valeur numérique - La satisfaction par rapport l'environnement
Genre	(1=FEMME, 2=HOMME)
Salaire Horaire	Valeur numérique - Salaire Horaire
Implication	Valeur numérique - L'implication de l'employé
Niveau hiérarchique	Valeur numérique
Le Poste	(1=Représentant médical, 2=RH, 3=Technicien Labo, 4=Manager, 5= Directeur de production, 6= Directeur de Recherche, 7= Chercheur, 8= Directeur des ventes, 9= Représentant commercial)
Satisfaction par rapport au poste	Valeur numérique - Satisfaction par rapport au poste
La situation familiale	(1=Divorcé, 2=Marié, 3=Célibataire)
Le revenu mensuel	Valeur numérique - Salaire Mensuel
Salaire Mensuel	Valeur numérique
Nombre d'ancien employeur	Valeur numérique
Sup à 18	(1=Oui, 2=No)
Heures supplémentaires	(1=Non, 2=Oui)
Augmentation de salaire	En pourcentage
Performance	Valeur numérique

Satisfaction relationnelle	Valeur numérique
Heures Normales	Valeur numérique
STOCK OPTIONS	Valeur numérique
Ancienneté Total	Valeur numérique
Nombre d'heures de formation	Valeur numérique
Équilibre vie privé/professionnelle	Valeur numérique
Ancienneté au sein de l'entreprise	Valeur numérique
Ancienneté dans le post actuel	Valeur numérique
Nombre d'année dès la dernière promotion	Valeur numérique
Nombre d'année avec le responsable actuel	Valeur numérique

*Source : Auteurs*