
HumanForYou

|Éthique|

RO11X202 - Projet IA - Groupe 4

Arthur Lecras, Nathan Poret, Valentin Pain, Benjamin Brifault



Table des matières

| | |
|---|----|
| Table d'illustrations | 2 |
| Introduction | 3 |
| Contexte | 3 |
| Objectifs de l'étude | 3 |
| Les problématiques générales de l'IA | 4 |
| La sélection des données | 5 |
| Liste des types et description des données disponibles | 5 |
| Filtrage éthique des données..... | 7 |
| Les droits fondamentaux..... | 7 |
| Question 1 : discrimination | 7 |
| Question 2 : droit des enfants | 7 |
| Question 3 : RGPD | 8 |
| Question 4 : liberté d'expression | 8 |
| Points de vigilances et données personnelles..... | 8 |
| Préparation des données | 9 |
| Interprétation éthique du modèle final et pistes de solutions | 10 |
| Conclusion | 11 |
| Bibliographie | 11 |

Table d'illustrations

| | |
|--|----|
| Figure 1 : Maintenir la mise en conformité des données avec le RGPD..... | 5 |
| Figure 2 : Le triangle CIA, une base essentielle de la sécurité | 10 |

Introduction

De nos jours, l'intelligence artificielle (IA) nous accompagne bien plus que nous aurions pu imaginer quelques années auparavant. Cette révolution s'est concrétisée dans de nombreux domaines et technologies allant de la simple utilité quotidienne jusqu'au secours d'une vie parfois. Toutefois, cette image de l'outil potentiellement parfait apporte son lot de contraintes lors de l'étape de conception. Entre les optimistes du transhumanisme et les sceptiques d'un ordinateur intelligent, nombreux sont ceux qui oublient les notions et les contraintes éthiques se trouvant au cœur de ces études.

Ce présent document constitue le premier livrable concernant le projet intelligence artificielle. Celui-ci se concentrera sur l'approche et la démarche éthique à adopter lors d'un projet d'une telle envergure (voir « Contexte »). Dans un premier temps, nous ferons un rappel relatif aux dangers et aux problématiques actuelles liées à l'intelligence artificielle et à la notion d'éthique. Enfin, nous aborderons notre projet concrètement ; Il est important de noter ici que ce document guidera l'ensemble de nos choix tout au long du projet, des justifications et une conclusion seront donc données sur les attentes et les résultats obtenus.

Contexte

L'entreprise HumanForYou subit depuis quelques années un turn-over d'environ 15% de ses employés tous les ans et nécessite de retrouver des profils similaires sur le marché de l'emploi. Celle-ci cherche à savoir les raisons de tous ces départs et a fait appel à notre équipe de spécialiste afin d'identifier ces différentes raisons. Ce turn-over n'est pas bon pour l'entreprise car les projets sur lesquels étaient les employés quittant la société prennent du retard ce qui nuit à la réputation de l'entreprise auprès de ses clients et partenaires. D'autre part, un service de ressources humaines de taille conséquente doit être conservé car il faut avoir les moyens de trouver les nouvelles recrues et du temps est perdu à l'arrivée des nouveaux employés car ils doivent très souvent être formés et ont besoin de temps pour devenir pleinement opérationnels dans leur nouvel environnement.

C'est dans ce contexte que cette entreprise de produits pharmaceutiques fait appel à nous. Nous avons à notre disposition certaines données nous permettant d'enquêter sur les raisons de ce turn-over annuel.

Objectifs de l'étude

D'un point de vue global du projet, notre but est de déterminer les facteurs ayant le plus d'influence sur ce taux de turn-over. La compréhension de ces facteurs nous permettra, par la suite, d'évoquer quelques pistes d'améliorations pour l'entreprise afin de réduire cet impact négatif annuel.

D'un point de vue plus éthique maintenant, nous devons faire une sélection des données sensibles parmi celles fournies par le service des ressources humaines. Cette sélection doit se faire avant tout développement et déterminera les axes d'apprentissage pour notre futur algorithme.

Cette étude se concentrera donc principalement sur de la conception et n'évoquera que très peu la partie technique.

Les problématiques générales de l'IA

L'intelligence artificielle offre aujourd'hui énormément de choses dans l'industrie, dans le médical ou plus simplement dans le quotidien de la population. Cependant, plusieurs questions se posent sur la fiabilité et la neutralité de ces systèmes.

Dans un premier temps, il est normal de se demander si l'intelligence artificielle ne remplacera pas totalement l'action humaine sur certaines tâches. En effet, les IA sont souvent estimées comme neutres par la plupart des utilisateurs qui ne sont pas sensibilisés à l'éthique alors qu'en réalité les choix de chaque IA reflètent les choix du développeur.

Cela nous amène donc à réfléchir sur le second point qui est le « biais discriminatoire ». Le biais discriminatoire est le fait de critiquer et de caractériser des personnes ou une population sur des aspects physiques ou liés à l'origine culturelle (genre, ethnies, sexualité...). Ce type de biais peut être volontaire ou involontaire selon les intentions des personnes à l'origine de l'IA.

La dernière problématique majeure de l'IA est la qualité ou la pertinence des données, mais aussi la quantité. Ces facteurs sont très importants dans le développement et l'alimentation d'une IA. C'est pourquoi, il est difficile de choisir un critère au détriment d'un autre. En réalité, si la quantité de donnée est privilégiée, cela peut être problématique, car l'IA n'aura potentiellement que des données de mauvaise qualité et sera alors moins efficace lors de sa mise en production. En revanche, si la qualité des données est optimale, l'IA pourra être plus précise et saura exactement ce que l'utilisateur recherche précisément. En conclusion, il est donc préférable de favoriser la qualité des données au détriment de la quantité pour développer une intelligence artificielle efficace et précise.

Il existe différentes méthodes pour éviter de faire face à ces problématiques. D'après les recommandations de la Commission européenne, une intelligence artificielle doit intégrer 7 exigences différentes :

1. Respect de l'autonomie humaine
2. Robustesse technique et sécurité
3. Confidentialité et gouvernance des données
4. Transparence
5. Diversité, non-discrimination et équité
6. Bien-être environnemental et sociétal
7. Responsabilité

De plus, le célèbre RGPD (Règlement général sur la protection des données), texte européen, établit un certain nombre de règles concernant la collecte et l'utilisation de données notamment à caractère personnel (on parle aussi de DCP : donnée à caractère personnel ou PII en anglais). Beaucoup de textes officiels et actuels se basent désormais sur cette réglementation.



Figure 1 : Maintenir la mise en conformité des données avec le RGPD

Celle-ci instaure notamment des sanctions telles que la suspension des données, des traitements ou encore la prononciation d'une amende (pouvant s'élever jusqu'à 20 millions d'euros) visant à dissuader l'exploitation non éthique des données et la recherche avide de performance.

La sélection des données

Liste des types et description des données disponibles

Comme évoqué plus haut, c'est le service de ressources humaines qui nous a fourni les fichiers de données.

Ces fichiers représentent chacun une partie bien spécifique :

- « **general_data.csv** » : les données générales sur les employés
- « **manager_survey_data.csv** » : les dernières évaluations du manager des employés
- « **employee_survey_data.csv** » : les résultats de l'enquête qualité de vie au travail pour chaque employé
- « **in_out_time.zip** » : archive contenant les informations sur les horaires de travail des employés. Celle-ci contient un fichier « in_time » contenant les horaires d'entrée chaque jour pour chaque employée et un fichier « out_time » pour les horaires de sortie

Contenu du fichier « general_data.csv » :

- « **Age** » : l'âge de l'employé en 2015
- « **Attrition** » : l'objet de notre étude, est-ce que l'employé a quitté l'entreprise durant l'année 2016 ?
- « **BusinessTravel** » : à quelle fréquence l'employé a été amené à se déplacer dans le cadre de son travail en 2015 (Non-Travel = jamais, Travel_Rarely = rarement, Travel_Frequently = fréquemment)
- « **DistanceFromHome** » : distance en km entre le logement de l'employé et l'entreprise
- « **Education** » : niveau d'étude : 1 = Avant College (équivalent niveau Bac), 2 = College (équivalent Bac+2), 3 = Bachelor (Bac+3), 4 = Master (Bac+5) et 5 = PhD (Thèse de doctorat)
- « **EducationField** » : domaine d'étude, matière principale

- « **EmployeeCount** » : booléen à 1 si l'employé était compté dans les effectifs en 2015
- « **EmployeeId** » : l'identifiant d'un employé
- « **Gender** » : sexe de l'employé
- « **JobLevel** » : niveau hiérarchique dans l'entreprise de 1 à 5
- « **JobRole** » : métier de l'employé dans l'entreprise
- « **MaritalStatus** » : statut marital de l'employé (Célibataire, Marié ou Divorcé)
- « **MonthlyIncome** » : salaire brut en roupies par mois
- « **NumCompaniesWorked** » : nombre d'entreprises pour lequel le salarié a travaillé avant de rejoindre HumanForYou
- « **Over18** » : est-ce que le salarié a plus de 18 ans ou non ?
- « **PercentSalaryHike** » : pourcentage d'augmentation du salaire en 2015
- « **StandardHours** » : nombre d'heures par jour dans le contrat du salarié
- « **StockOptionLevel** » : niveau d'investissement en actions de l'entreprise par le salarié
- « **TotalWorkingYears** » : nombre d'années d'expérience en entreprise du salarié pour le même type de poste
- « **TrainingTimesLastYear** » : nombre de jours de formation en 2015
- « **YearsAtCompany** » : ancienneté dans l'entreprise
- « **YearsSinceLastPromotion** » : nombre d'années depuis la dernière augmentation individuelle
- « **YearsWithCurrentManager** » : nombre d'années de collaboration sous la responsabilité du manager actuel de l'employé

Contenu du fichier « manager_survey_data.csv » :

- « **EmployeeID** » : L'identifiant de l'employé
- « **JobInvolvement** » : Une évaluation de son implication dans son travail notée 1 ('Faible'), 2 ('Moyenne'), 3 ('Importante') ou 4 ('Très importante')
- « **PerformanceRating** » : Une évaluation de son niveau de performance annuel pour l'entreprise notée 1 ('Faible'), 2 ('Bon'), 3 ('Excellent') ou 4 ('Au-delà des attentes')

Contenu du fichier « employee_survey_data.csv » :

- « **EmployeeID** » : l'identifiant de l'employé
- « **EnvironmentSatisfaction** » : l'environnement de travail, noté 1 ("Faible"), 2 ("Moyen"), 3 ("Élevé") ou 4 ("Très élevé")
- « **JobSatisfaction** » : son travail, noté de 1 à 4 comme précédemment
- « **WorkLifeBalance** » : son équilibre entre vie professionnelle et vie privée, noté 1 ("Mauvais"), 2 ("Satisfaisant"), 3 ("Très satisfaisant") ou 4 ("Excellent")

Contenu des fichiers « in_time.csv » et « out_time.csv » :

- « **EmployeeID** » : l'identifiant de l'employé
- « **Les horaires d'arrivée** » (pour le fichier in_time)
- « **Les horaires de départ** » (pour le fichier out_time)

Filtrage éthique des données

Comme il est mentionné dans la « Liste d'évaluation pour une IA digne de confiance » de la Commission européenne. Les données utilisées ne doivent pas discriminer négativement le choix de l'IA. Il faut pour cela éliminer les données liées au sexe, à la couleur de peau, à l'ethnie, à l'origine sociale, au statut social, au statut marital, à la langue, à la religion, aux choix politiques ou toutes autres opinions personnelles, à l'âge, à la date ou le lieu de naissance, à l'appartenance à une minorité culturelle et à l'orientation sexuelle...

A l'aide de ces directives nous éliminons donc les colonnes suivantes de notre jeu de données :

- « **Gender** » : représente le genre de l'employé
- « **MaritalStatus** » : représente un statut marital de l'employé
- « **Age** » : représente l'âge de l'employé
- « **Over18** » : représente le statut social et l'âge de l'employé

Les droits fondamentaux

Le concept de droits fondamentaux dans le domaine de l'intelligence artificielle est mis en avant et est très apprécié des utilisateurs d'IA. Ce concept pose 4 questions primordiales permettant de savoir si une intelligence artificielle est bien éthique. Ces questions sont centrées sur le droit humain ainsi que le respect de la vie privée et la protection des données sensible et personnelles.

Pour vérifier que notre système de Machine Learning est en accord avec les droits fondamentaux, nous allons répondre à ces questions les unes après les autres.

Question 1 : discrimination

Question : Le système d'IA est-il potentiellement discriminatoire à l'égard des personnes sur la base de l'un des motifs suivants (de manière non exhaustive) : sexe, race, couleur, origines ethniques ou sociales, caractéristiques génétiques, langue, religion ou croyance, opinions politiques ou autres, appartenance à une minorité nationale, propriété, naissance, handicap, âge ou orientation sexuelle ?

Réponse : Certains types de données (voir la partie « filtrage éthique des données » plus haut) présents dans la liste originale fournie par l'entreprise HumanForYou ont rapidement été écartés car jugés non éthiques par notre équipe et au regard des textes de la Commission européenne.

Question 2 : droit des enfants

Question : Le système d'IA respecte-t-il les droits de l'enfant, par exemple en matière de protection de l'enfance et de prise en compte de l'intérêt supérieur de l'enfant ?

Réponse : Aucune donnée présente dans notre ensemble de données ne vise à représenter une caractéristique directe et personnelle liée à un enfant mineur (- 18 ans).

Question 3 : RGPD

Question : Le système d'IA protège-t-il les données personnelles relatives aux individus conformément au RGPD ?

Réponse : Toutes les données liées aux employés ne permettent pas d'identifier un individu de l'entreprise en particulier. La plupart des données catégorielles ont été numérisées et chaque employé possède un identifiant remplaçant ses informations personnelles.

Question 4 : liberté d'expression

Question : Le système d'IA respecte-t-il la liberté d'expression et d'information et/ou la liberté de réunion et d'association ?

Réponse : L'objectif de cette IA est de déterminer les raisons du turn-over annuel d'employés dans l'entreprise et par extension de déterminer comment l'entreprise peut améliorer les conditions de travail des employés. C'est justement de cette manière que celle-ci pourra potentiellement prendre des décisions concernant l'environnement de travail global et ainsi l'épanouissement et la liberté d'expression de ses employés.

Points de vigilances et données personnelles

La « Liste d'évaluation pour une IA digne de confiance » énonce que le système doit impérativement protéger les données personnelles relatives aux individus. Ces données sensibles (PII) doivent donc être identifiées dans le jeu de données pour être anonymisées et donc protéger l'individu de toutes discriminations qu'elles soient positives ou négatives. Le RGPD indique que toutes les données servent bien un but précis. Autrement dit, il ne faut pas collecter les informations simplement dans le cas où celles-ci seront utiles un jour. À noter également que cette partie ne parlera pas des données déjà écartées dans la partie « filtrage éthique des données ».

En rassemblant les différentes données personnelles possible, les colonnes contenant des données personnelles sont les suivantes :

- « **DistanceFromHome** » : contient la localisation approximative de l'employé
- « **Education** » : contient le niveau d'étude de l'employé
- « **EducationField** » : contient le secteur d'étude de l'employé
- « **WorkLifeBalance** » : contient une note de l'équilibre entre la vie privée et la vie personnelle de l'employé
- « **NumCompaniesWorked** » : contient le nombre d'entreprises pour lesquelles l'employé a travaillé précédemment
- « **PercentSalaryHike** » : contient le pourcentage d'augmentation du salaire de l'employé

- « **StockOptionLevel** » : contient le nombre d'actions de l'employé dans l'entreprise
- « **TotalWorkingYears** » : contient le nombre d'années travaillées par l'employé
- « **YearsSinceLastPromotion** » : contient l'année depuis la dernière promotion
- « **JobRole** » : contient le métier de l'employé
- « **BusinessTravel** » : contient la fréquence de voyage d'affaires de l'employé
- « **TrainingTimesLastYear** » : contient le nombre de jours de formation dans l'année de l'employé
- « **YearsAtCompany** » : contient l'ancienneté de l'employé
- « **Les heures de pointages** » : contient toutes les heures d'entrée et de sortie d'un employé par jours

Remarque : Comme précisé plus haut dans ce document, chaque employé a été anonymisé à l'aide d'un identifiant numérique unique (EmployeeId).

Préparation des données

L'étape de la préparation des données est très importante lorsque qu'on fait de l'IA car c'est elle qui va influencer le plus sur la qualité des résultats et surtout sur la nature des données. Une bonne préparation des données passe par un filtrage éthique (avant entraînement des modèles) et une série de jointures, nettoyages, suppressions, mises à l'échelle, normalisations et standardisations.

Pour commencer, les fichiers « in_time » et « out_time » ont été fusionnés afin d'obtenir un seul fichier avec tous les horaires. Une fois les horaires standardisés, une nouvelle colonne va voir le jour : « *AttendanceScore* », cette colonne décrit, avec un score, le respect des horaires de l'employé dans l'année 2015. L'individu subira une pénalité s'il ne respecte pas les horaires de son contrat, un bonus s'il fait des heures supplémentaires et rien s'ils sont respectés. Celle-ci prenant en compte la somme du respect des horaires, cela va donc permettre de réduire le nombre de colonnes total du jeu de données. Enfin cette colonne « *AttendanceScore* » a été ajoutée au fichier dans lequel les jeux de données « general_data », « employee_survey_data » et « manager_survey_data » ont été fusionnés au préalable. Comme il n'existe qu'une seule valeur pour l'ensemble des employés, la colonne « EmployeeId » a elle aussi été supprimée. Les données sont donc maintenant complètement anonymisées.

Ensuite, les valeurs non numériques ont été encodées. Par exemple, la colonne au cœur du problème, « Attrition » met en avant si un employé a quitté ou non l'entreprise en 2016. Celle-ci s'est vu remplacer ses « Yes » par des « 1 » et ses « No » par des « -1 ».

Enfin, une fois les données mises en forme, il s'agissait de réduire les critères envoyés dans l'IA. Pour cela, une matrice de corrélation a été réalisée entre les différentes colonnes et la colonne « Attrition ». Après la mise en évidence des valeurs obtenues, de nombreuses colonnes ont été supprimées ne laissant à la fin seulement les cinq colonnes suivantes :

- « **TotalWorkingYear** » : nombre d'années d'expérience en entreprise du salarié pour le même type de poste
- « **YearsAtCompany** » : ancienneté dans l'entreprise

- « **YearsWithCurrManager** » : nombre d'années de collaboration sous la responsabilité du manager actuel de l'employé
- « **EnvironmentSatisfaction** » l'environnement de travail, noté 1 ("Faible"), 2 ("Moyen"), 3 ("Élevé") ou 4 ("Très élevé")
- « **AttendanceScore** » : score de présence de l'employé sur l'année

Les données sont désormais bien préparées et respectent l'anonymat de chaque employé tout en préservant la cohérence des données nécessaire afin que l'IA fournisse un résultat exploitable.

Interprétation éthique du modèle final et pistes de solutions

Après la sélection du meilleur modèle en fonction de ses métriques et des besoins du client, il sera possible d'exploiter le résultat de celui-ci. Cependant, il ne faut pas négliger la manière d'interpréter ces données. Après entraînement, le modèle final choisi est le « DecisionTreeClassifier » et présente les colonnes « *TotalWorkingYear* » (nombre d'années d'expérience en entreprise du salarié pour le même type de poste) et « *AttendanceScore* » comme étant les données agissant le plus sur ce turn-over annuel. La colonne « *AttendanceScore* » représentant le score d'assiduité, il est aussi à conclure que les employés ne respectent pas vraiment les horaires imposés.

Néanmoins, il ne faut pas pour autant agir en les sanctionnant ou en les pénalisant d'une façon ou d'une autre. Il faut au contraire donner envie à l'employé de rester, par exemple en diversifiant les missions sur son poste, en lui donnant plus de responsabilités ou en réorientant la personne. Cela permettra de supprimer les coutumes ennuyeuses et répétitives et ainsi susciter de l'intérêt. Offrir une « récompense » est aussi très conseillé pour pousser ceux-ci à vouloir faire plus et donc rester dans l'équipe. L'entreprise se doit d'ailleurs de réaliser plus de sondages sur le bien-être de ses salariés dans leur environnement de travail ou encore mettre en place des campagnes de sensibilisation sur ce sujet.

En dernier lieu, et afin de garantir un respect complet de la mise en place de cette démarche, l'ensemble des données se doit d'être correctement sécurisé aussi bien physiquement que logiquement. Celles-ci ne doivent être accessibles que par les personnes de l'entreprise autorisées et qualifiées en fonction du dispositif de sécurité prévu à cet effet.



Figure 2 : Le triangle CIA, une base essentielle de la sécurité

Il est à rappeler également que le traitement d'une information ne se fait pas obligatoirement de façon numérique et informatisé. Le même soin pour la protection des données informatisées doit être appliqué pour les données au format papier.

Conclusion

L'intelligence artificielle est de plus en plus utilisée aujourd'hui par les entreprises pour un grand nombre de besoins spécifiques. Cette technologie est effectivement très intéressante et permet de gagner beaucoup de temps sur de nombreuses tâches chronophages. Néanmoins, il ne faut pas oublier que celle-ci est également très controversée, en particulier lorsqu'il s'agit de manipuler et de traiter des données sensibles et/ou personnelles. C'est pour cette raison que certaines mesures ont été prises pour limiter les biais des IA en rapport à l'éthique avec notamment l'anonymisation des données sensibles du personnel ou encore la mise à l'écart des données non éthiques et inutiles pour l'étude.

De plus, dans notre cas, cette étude permettra d'implémenter de nouvelles solutions destinées à améliorer la vie au travail des salariés et non d'obtenir des informations à des fins non définies. Il ne faut également pas oublier que l'intelligence artificielle n'est qu'un outil d'information et non de décision. L'humain doit être et rester le décisionnaire final dans tous les cas.

Bibliographie

Éthique

- « Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment », s. d., <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>.
- « Éthique et IA : analyse et discussion », s. d. <https://hal.archives-ouvertes.fr/hal-03321149/document>.
- « Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle », s. d. <https://www.cnil.fr/fr/comment-permettre-lhomme-de-garder-la-main-rapport-sur-les-enjeux-ethiques-des-algorithmes-et-de>.

RGPD

- « Comment gérer les données sensibles RGPD », s. d. <https://donneespersonnelles.fr/comment-gerer-les-donnees-sensibles-rgpd>.
- « Comment (ré)concilier RGPD et big data ? », s. d. <https://pure.unamur.be/ws/files/54216394/8327.pdf>.
- « RGPD : de quoi parle-t-on ? », s. d. <https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on>.
- « Qu'est-ce qu'une donnée sensible RGPD ? », s. d. <https://donneespersonnelles.fr/donnee-sensible-rgpd>.

Sanction en cas de non-respect de l'éthique

- « Sanction éthique - RGPD », s. d. <https://www.cnil.fr/fr/definition/sanction>.



Sécurité des données

« Les politiques de sécurité », s. d.

<http://www2.ift.ulaval.ca/~desharnais/Recherche/RR/TherrezinhaPolitiques2001.pdf>.