

Projet

[Objectifs d'apprentissage](#)[Ressources pour les étudiants](#)

Sujet

IA for HumanForYou

L'entreprise de produits pharmaceutiques HumanForYou basée en Inde emploie environ 4000 personnes. Cependant, chaque année elle subit un turn-over d'environ 15% de ses employés nécessitant de retrouver des profils similaires sur le marché de l'emploi.

La direction trouve que ce niveau de 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.
- Un service de ressources humaines de taille conséquente doit être conservé car il faut avoir les moyens de trouver les nouvelles recrues.
- 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.

Le direction fait donc appel à vous, spécialistes de l'analyse de données, pour déterminer les facteurs ayant le plus d'influence sur ce taux de turn-over et lui proposer des modèles afin d'avoir des pistes d'amélioration pour donner à leurs employés l'envie de rester.

Données fournies

Un certain nombre de données concernant les employés vous a donc été transmis par le service des ressources humaines.

Il s'agit de fichiers textes au format CSV.

Les données ont été anonymisées : un employé de l'entreprise sera représenté par le même EmployeeID dans l'ensemble des fichiers qui suivent.

Données du service des ressources humaines

Pour chaque employé, le service des ressources humaines vous confie les informations en sa possession :

- 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 : A quel 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 dans l'entreprise
- MaritalStatus : Statut marital du salarié (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 : % 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.

general_data.csv [csv] ↕

Dernière évaluation du manager

Ce fichier contient la dernière évaluation de chaque employé faite par son manager en février 2015.

Il contient les données suivantes :

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

manager_survey_data.csv [csv] ↕

Enquête qualité de vie au travail

Ce fichier provient d'une enquête soumise aux employés en juin 2015 par le service RH pour avoir un retour concernant leur qualité de vie au travail.

Une organisation avait été mise en place pour que chacun puisse répondre à ce questionnaire sur son lieu de travail en concertation avec les managers mais il n'y avait pas d'obligation.

Les employés devaient répondre à 3 questions sur le niveau de satisfaction concernant :

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

Lorsque un employé n'a pas répondu à une question, le texte "NA" apparaît à la place de la note.

employee_survey_data.csv [csv] ↕

Horaires de travail

Des badgeuses sont installées et utilisées dans l'entreprise depuis quelques années. Il a été jugé opportun par la direction de vous transmettre les horaires d'entrée et de sortie des employés sur une période de l'année choisie représentative d'une activité moyenne pour l'ensemble des services.

Vous trouverez donc 2 fichiers traçants les horaires d'arrivée à leur poste et de départ de leur poste de l'ensemble des employés par date sur une période allant du 1^{er} janvier au 31 décembre 2015.

in_out_time.zip [zip] ↕

Sources :

Les données de ce projet sont des données générées issues de ce projet sur Kaggle :

<https://www.kaggle.com/vjchoudhary7/hr-analytics-case-study> ↗


Vous êtes attendu sur votre capacité à exploiter les données fournies pour répondre aux besoins de l'entreprise qui vous a sollicité. Vous pouvez vous inspirer d'approches faites par d'autres personnes mais vous devrez être en mesure d'expliquer chacun des choix retenus en les justifiant.

▲ Livrables attendus

Projet I.A. - Livrable : éthique

Il s'agit dans un document mettant en évidence votre appropriation de la méthodologie proposée pour avoir une démarche éthique traçant vos choix justifiés par rapport aussi aux échanges qu'il y a eu dans votre équipe pendant

toutes les phases du projet de la préparation du jeu de données à la proposition de pistes d'amélioration à votre client. En plus de la démarche, vous devrez bien mettre en évidence les décisions que vous avez prises dans l'équipe ainsi que les points de vigilance ou de contrôle à mettre en place par rapport aux problématiques éthiques rencontrées.

Pour vous guider dans cette démarche, vous pouvez vous appuyer les questions accessibles depuis ce site . Le fichier PDF téléchargeable donne les principales questions à se poser pour valider qu'un projet d'IA est éthique selon 7 exigences recommandées par la Commission Européenne :

- respect de l'autonomie humaine
- robustesse technique et sécurité
- confidentialité et gouvernance des données
- transparence
- diversité, non-discrimination et équité
- bien-être environnemental et sociétal
- responsabilité

Projet I.A. - Présentation

Il s'agit de votre présentation intégrant votre notebook Jupyter (qui sera envoyé la veille de la soutenance) et votre présentation de 20 minutes qui devra montrer l'ensemble de votre démarche :

- La génération de votre/vos jeux de données en justifiant les choix et les traitements effectués pour faire face aux problématiques classiques de ce type de travail.
- le choix de (des) l'algorithme(s) d'IA retenu(s).
- l'analyse des résultats obtenus et leur interprétation en s'appuyant sur des métriques.
- La démarche mise en œuvre pour améliorer un modèle.
- le choix du modèle retenu au final parmi les différents essais effectués en justifiant pas rapport au besoin de votre client et des métriques.
- vos propositions finales justifiées

