

# PROJET

Tout afficher + Tout réduire -

## SUJET

### ✓ IA for HumanForYou

#### IA for HumanForYou

L'entreprise pharmaceutique HumanForYou, basée en Inde, emploie environ 4000 personnes. Cependant, chaque année, elle connaît un taux de rotation d'environ 15 % de ses employés, ce qui nécessite de trouver des profils similaires sur le marché de l'emploi.

La direction considère que ce niveau de rotation n'est pas bénéfique pour l'entreprise pour les raisons suivantes :

- Les projets sur lesquels travaillaient les employés quittant la société prennent du retard, ce qui nuit à la réputation de l'entreprise auprès de ses clients et partenaires.
- Il est nécessaire de conserver un service des ressources humaines conséquent afin d'avoir les moyens de trouver de nouveaux employés.
- L'arrivée des nouveaux employés entraîne une perte de temps, car ils doivent souvent être formés et ont besoin d'un certain temps pour être pleinement opérationnels dans leur nouvel environnement.

La 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 rotation et pour proposer des modèles afin d'identifier des pistes d'amélioration permettant de motiver les employés à rester dans l'entreprise.

#### Données fournies

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

Il s'agit de fichiers texte au format CSV.

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

#### Données du service des ressources humaines

#### 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, 550,0 ko] ⬇

### Dernière évaluation du manager

#### 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, 43,0 ko] ⬇

### Enquête qualité de vie au travail

#### 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, 52,0 ko] ⬇

### Horaires de travail

#### 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 qui retracent les horaires d'arrivée à leur poste et de départ de leur poste pour l'ensemble des

employés par date sur une période allant du 1er 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 sollicitée. 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 d'un document mettant en évidence votre appropriation de la méthodologie proposée pour adopter une démarche éthique en justifiant vos choix, en tenant compte des échanges au sein de votre équipe tout au long du projet, depuis la préparation des données jusqu'à la proposition de pistes d'amélioration à votre client. En plus de la démarche, vous devrez mettre en évidence les décisions prises en équipe ainsi que les points de vigilance ou de contrôle à mettre en place concernant les problématiques éthiques rencontrées.

Pour vous guider dans cette démarche, vous pouvez vous appuyer les questions accessibles depuis ce site 📄. Le fichier PDF téléchargeable contient les principales questions à se poser pour évaluer l'éthique d'un projet d'IA, selon les 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. - Livrable : Bibliographie

Il s'agit d'un document visant à présenter les références académiques et techniques que vous avez utilisées pour orienter votre travail, en indiquant leur pertinence dans la compréhension des concepts abordés. Ce document inclut des sources permettant de justifier vos choix, de valider les méthodes employées et de guider les prises de décision dans le cadre de votre projet. En plus de citer les ouvrages et articles, il est essentiel de souligner les apports de chaque source ainsi que leur impact sur le développement de votre démarche.

Pour organiser votre bibliographie, vous pouvez classer les sources par thématique :

- **Sources méthodologiques et théoriques** : références traitant des bases théoriques et des modèles méthodologiques employés.
- **Sources sur les aspects techniques** : ouvrages ou articles décrivant les techniques spécifiques utilisées.
- **Sources éthiques et sociétales** : ressources relatives aux normes et aux enjeux éthiques dans votre domaine.
- **Sources spécifiques au projet** : articles, études de cas ou rapports ayant directement inspiré ou orienté vos choix.

Pour chaque source, veillez à inclure :

- Le respect des droits d'auteur et des licences de publication.
- Une brève annotation justifiant l'intégration de chaque source dans votre bibliographie.

### 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 l'algorithme ou des algorithmes d'IA sélectionnés
- L'analyse des résultats obtenus et leur interprétation en se basant 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

