

Analyse des ressources humaines

Partie 1

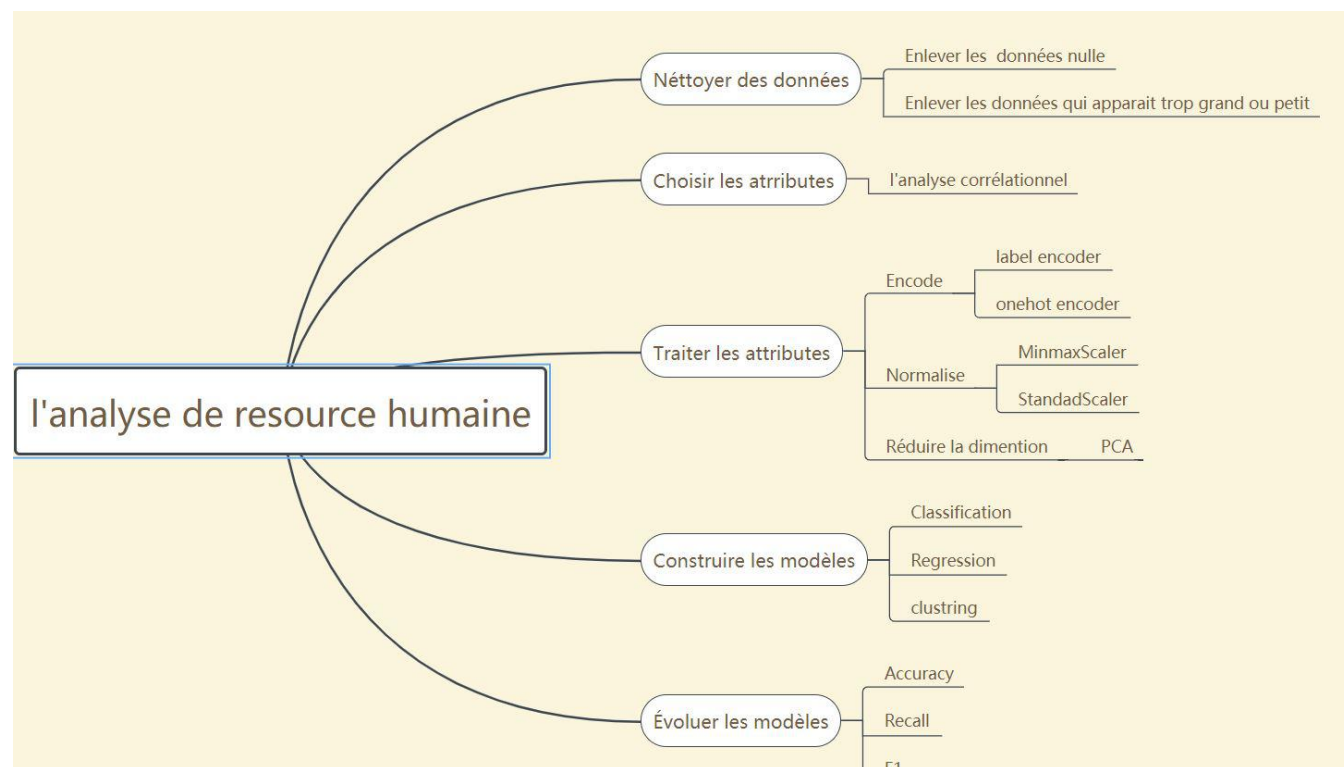
Présentation

Objective:

Pour comprendre les raisons pour les quelle les employées ont démissionnées et créer un modèle pour prédire si un emploé va quitter l'entreprise. L'objectif est d'améliorer les stratégies pour retenir les employées. L'implémentation de modèle va permettre de faciliter le management et faire de meilleur décisions

Partie 2

Schéma de l'exploration



Partie 3

Exploration des données

3.a Explorer les données ---- Choisir les attributs:

- Heatmap

1. La relation positive:

Number project VS last_evaluation

Number project VS average_monthly_hours

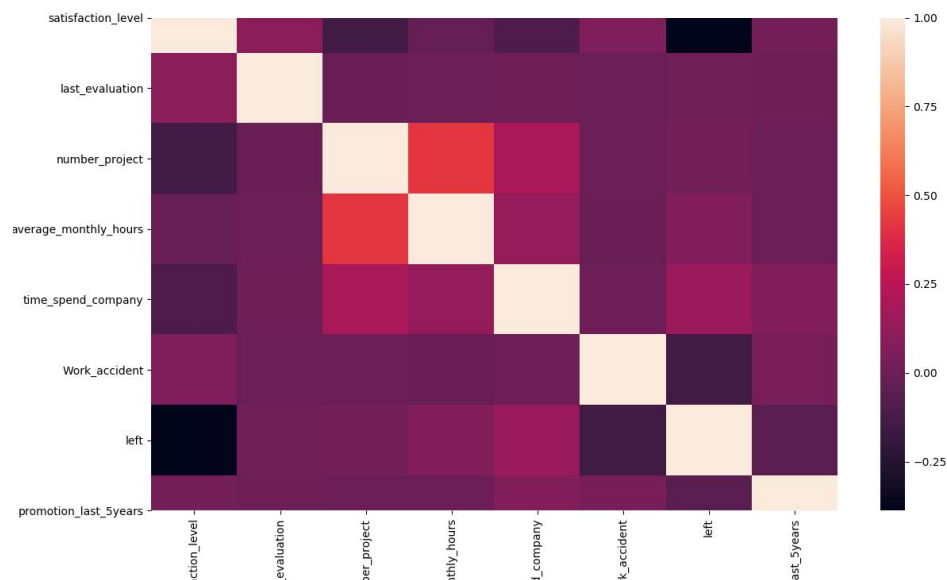
- ## 2. La relation négative

Satisfaction_level Vs last_evaluation

- La résumé

Selon le heatmap, on peut constater qu'il y a une relation positive entre le nombre de projet, et average hours et evaluation. Ca veut dire que les employés qui dépense plus de temps et faire plus de projects evaluent plus vite.

Pour la relation négative, les employées démissionnée et la satisfaction est liée, j'assume que les employées qui ont une tendance de partir de l'entreprise quand ils sont moins satisfaits.



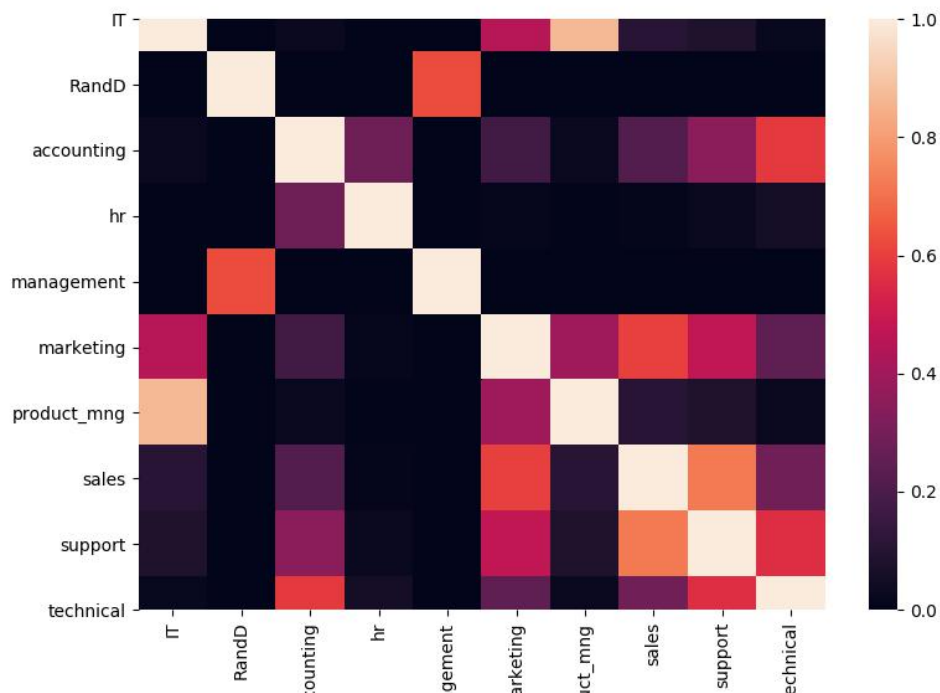
Heapmap2 :

La relation entre les département différents

- La relation positive:

Département IT Vs Département produits

Pour l'exploration plutard, on peut remarquer qu'il y a une relation lié entre les actifs de la département en informatique et les actifs en produits.



3.b Distribution Plots(Satisfaction - Evaluation - AverageMonthlyHours)

- La résumé:

Satisfaction:

il y a une grande écart entre la satisfaction haut et la satisfaction bas

Évaluation:

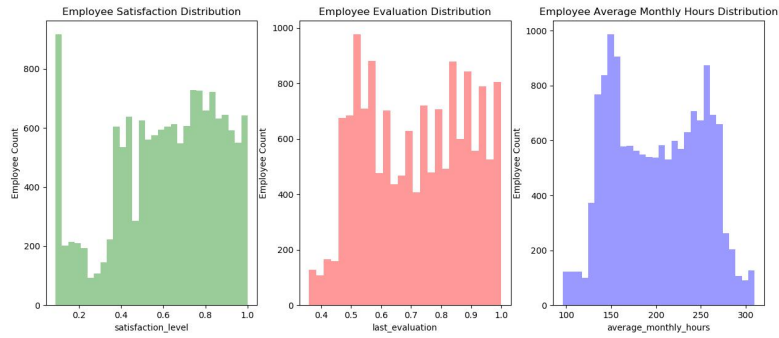
La distrubtion de l'évaluation est réparti entre moin que 0.6 et plus que 0.8

AverageMonthlyHours:

La distribution entre l'évaluation et l'avearge heures se ressemble, ca veut dire que les employée qui travaille moins ont moins de l'évaluation et vice versa, on peut aussi remarquer ce résultat selon le heatmap

- Question à poser:

Pourquoi il y a une sommet pour les employées avec satisfaction bas



3.c Salary VS Left

- La résumée:

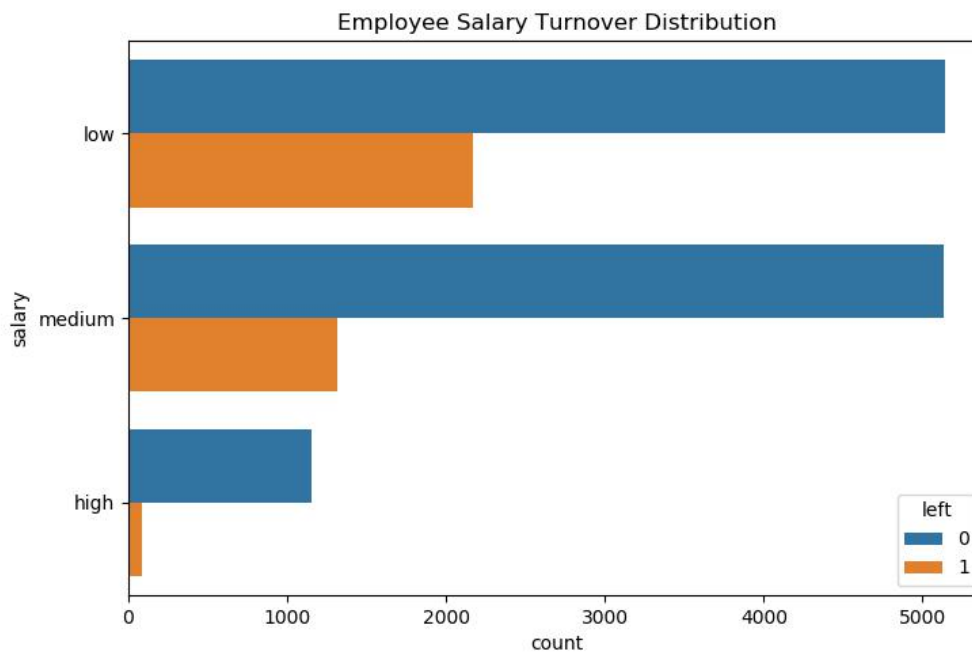
La plupart de gens qui ont quitté leur travail ont un salaire bas ou moyen

Rarement les employées avec un salaire haut démissionne

Mais , il y a aussi les employées qui quitte avec le salaire haut,

- Question à poser :

Pour quelle rais que les employées avec la salaire haut quittent l'entreprise

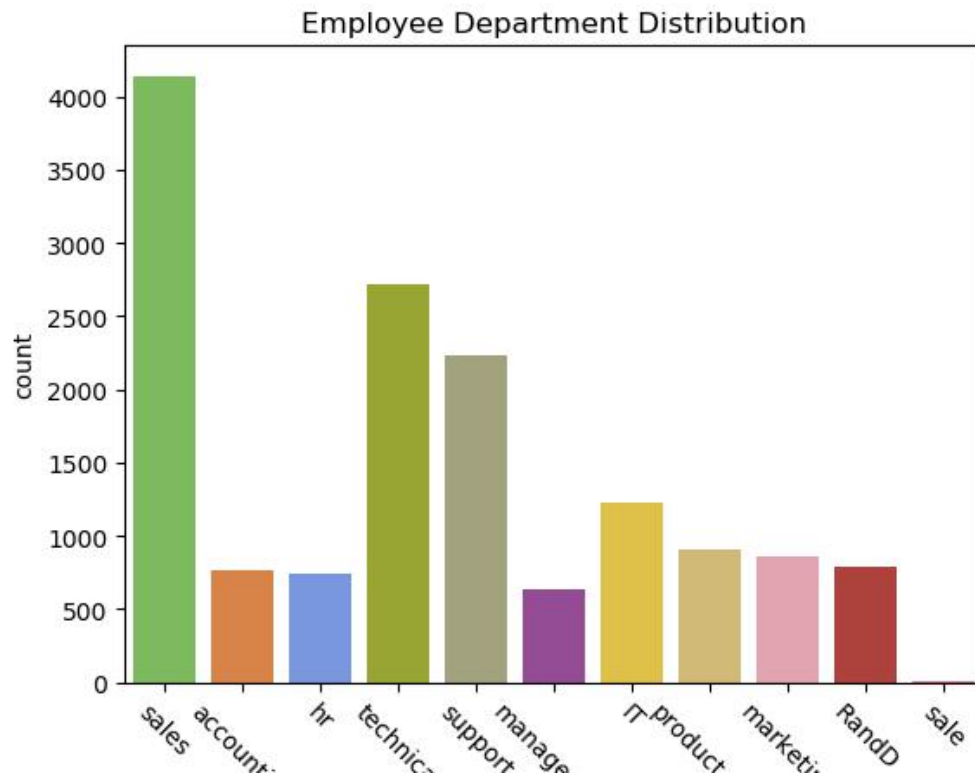


3.d département vs left

- La résumée:

Le département sales, technical, and le support de département sont les 3 premières à quitter l'entreprise)

Le département de management a le moins de possibilité de quitter

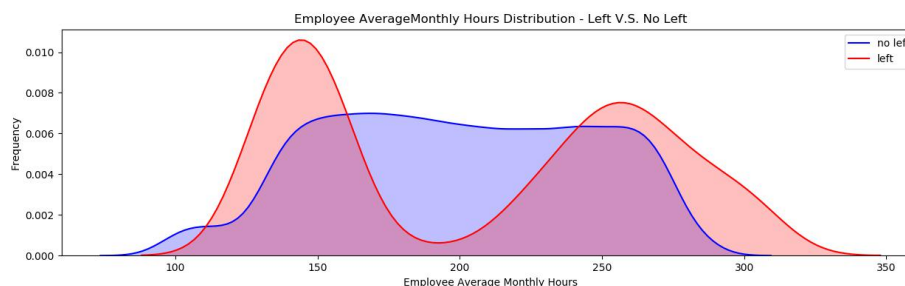


3.e Left vs Average hours

- La résumée

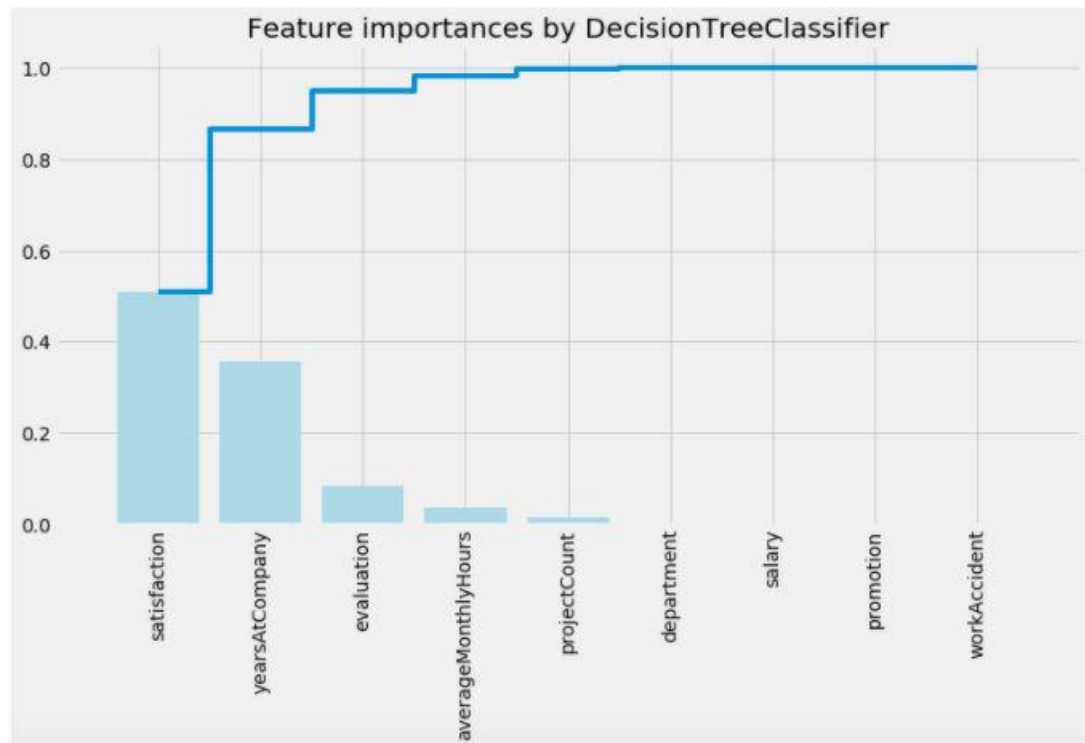
Les employees qui font moins de travailles (moins que 150heurs) quittent l'entreprise

Les employés qui font trop d'heures de travail ont aussi plus de chance de quitter l'entreprise



4.a Modèles :

Décision tree pour avoir des attributs les plus important afin de construire le modèle logistic regression, c'est plus interprétable pour comprendre



4.b logistic regression analyse

: Éliminer les autres variables, et analyser ces trois attributs principales, the constant 0.181896 représente the effet incontrôable

```
Optimization terminated successfully.  
Current function value: 0.467233  
Iterations 6
```

```
satisfaction      -3.769022  
evaluation         0.207596  
yearsAtCompany    0.170145  
int               0.181896  
dtype: float64
```

4.c pour les autres modèles: Random Forest , Adaboost ,
Logistic, Decision Tree

```
---Logistic Model---  
Logistic AUC = 0.74  
      precision    recall  f1-score   support  
  
0       0.90      0.76      0.82      1714  
1       0.48      0.73      0.58       536  
  
avg / total       0.80      0.75      0.76      2250
```

```
---Decision Tree Model---  
Decision Tree AUC = 0.94  
      precision    recall  f1-score   support  
  
0       0.97      0.96      0.97      1714  
1       0.87      0.91      0.89       536  
  
avg / total       0.95      0.95      0.95      2250
```

Random Forest AUC = 0.97

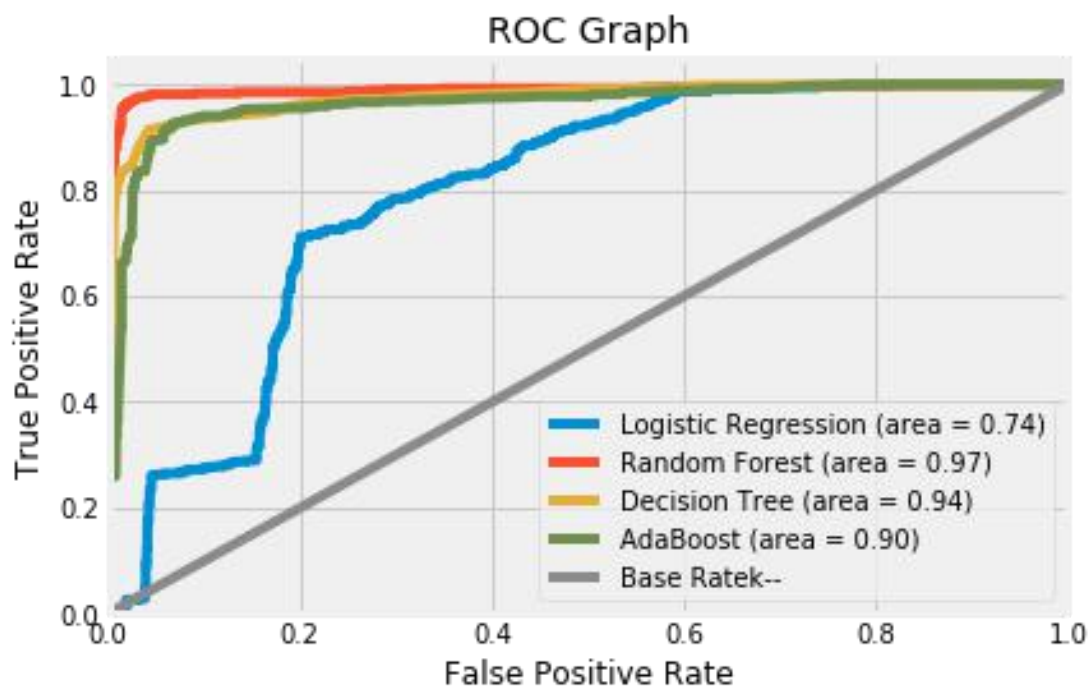
	precision	recall	f1-score	support
0	0.99	0.98	0.99	1714
1	0.95	0.96	0.95	536
avg / total	0.98	0.98	0.98	2250

---AdaBoost Model---

AdaBoost AUC = 0.90

	precision	recall	f1-score	support
0	0.95	0.97	0.96	1714
1	0.90	0.82	0.86	536
avg / total	0.93	0.94	0.93	2250

L'évaluation de ces modèles



5. Intéprèteter ces résultats:

- Les employées ont risque de quitter leur poste quand ils travaille pas assez heures (moins que 150hr/mois ou 6hr/jour), ou travaille très longtemps(plus que 250hr par mois ou 10 hr par jour)
- Les employées avec les salaire au mieu ou très bas vont quitter l'entreprise
- Les attributes qui influence le plus :

Satisfaction , les années rester en entreprise, l'évaluation

6. Application et objective atteinte:

Au lieu de simplement dire si un employée va quitter l'entreprise ou pas. On utilise la propabilité pour prédire. Donc on peut classer les employées par leur probabilité de démissionner et essayer de retenir les employées.