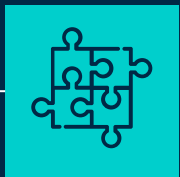


PROJET – ANALAYSE DES TURNOVER

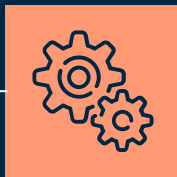
QUESNEL Ninon – BERTON Léonie – REVERSAC PAUL – PERICHON Nicolas

SOMMAIRE



01

JEU DE DONNEES



02

METHODOLOGIE



03

MODELES



04

RESULTATS



05

APPROFONDIR



06

CONCLUSION



01 JEU DE DONNEES

LIBELLE	TYPE	DESCRIPTION
Satisfaction	float	Note /1 représentant le taux de satisfaction du salarié dans son métier.
Derniere_evaluation	float	Dernière note attribuée au collaborateur.
Nombre_de_projets	int	Nombre de missions sur lesquelles le salarié travaille.
Nombre_heures_mensuelles_moyenne	int	Nombre d'heures travaillées en moyenne par mois pour un employé.
Temps_passe_dans_entreprise	int	Nombre d'années passées dans l'entreprise par le salarié.
Accident_du_travail	int	L'employé a déjà eu un accident du travail ou non.
Depart	int	L'employé a-t-il quitté l'entreprise ? (1 : Oui / 0 : Non)
Promotion_5_dernieres_annees	int	Le salarié a-t-il eu une promotion ces 5 dernières années ?
Service	object	Service dans lequel l'employé travaille.
Niveau_salaire	object	Niveau de salaire de l'employé (low,medium,high).



01 JEU DE DONNEES

- 14 950 enregistrements / 10 variables
- Def `resumetable(df)`
- Variable cible : départ (0 : Non / 1 : Oui)

	Name	dtypes	Missing	Uniques	First value	Second value	Third value
0	Satisfaction	float64	0	92	0.41	0.87	0.45
1	derniere_evaluation	float64	0	65	0.54	0.88	0.48
2	Nombre_de_projets	int64	0	6	2	5	2
3	Nombre_heures_mensuelles_moyenne	int64	0	215	152	269	158
4	Temps_passe_dans_entreprise	int64	0	8	3	5	3
5	Accident_du_travail	int64	0	2	0	0	0
6	depart	int64	0	2	1	1	1
7	promotion_5_dernieres_annees	int64	0	2	0	0	0
8	Service	object	0	10	technical	technical	technical
9	niveau_salaire	object	0	3	low	low	low



01 JEU DE DONNEES

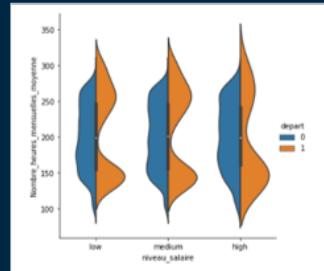
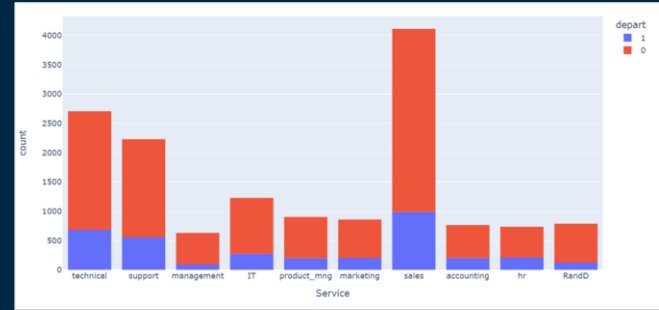
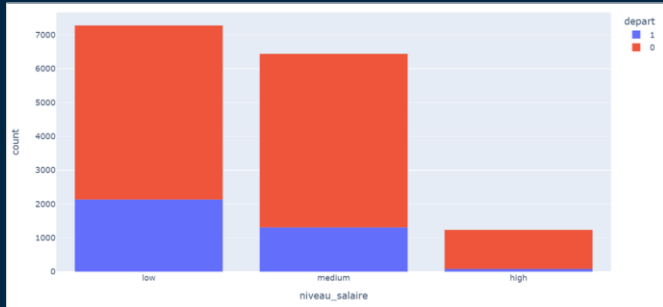
- OrdinalEncoder()

	Name	dtypes	Missing	Uniques	First value	Second value	Third value
0	Satisfaction	float64	0	92	0.41	0.87	0.45
1	derniere_evaluation	float64	0	65	0.54	0.88	0.48
2	Nombre_de_projets	int64	0	6	2.00	5.00	2.00
3	Nombre_heures_mensuelles_moyenne	int64	0	215	152.00	269.00	158.00
4	Temps_passe_dans_entreprise	int64	0	8	3.00	5.00	3.00
5	Accident_du_travail	int64	0	2	0.00	0.00	0.00
6	depart	int64	0	2	1.00	1.00	1.00
7	promotion_5_dernieres_annees	int64	0	2	0.00	0.00	0.00
8	Service	float64	0	10	9.00	9.00	9.00
9	niveau_salaire	float64	0	3	1.00	1.00	1.00



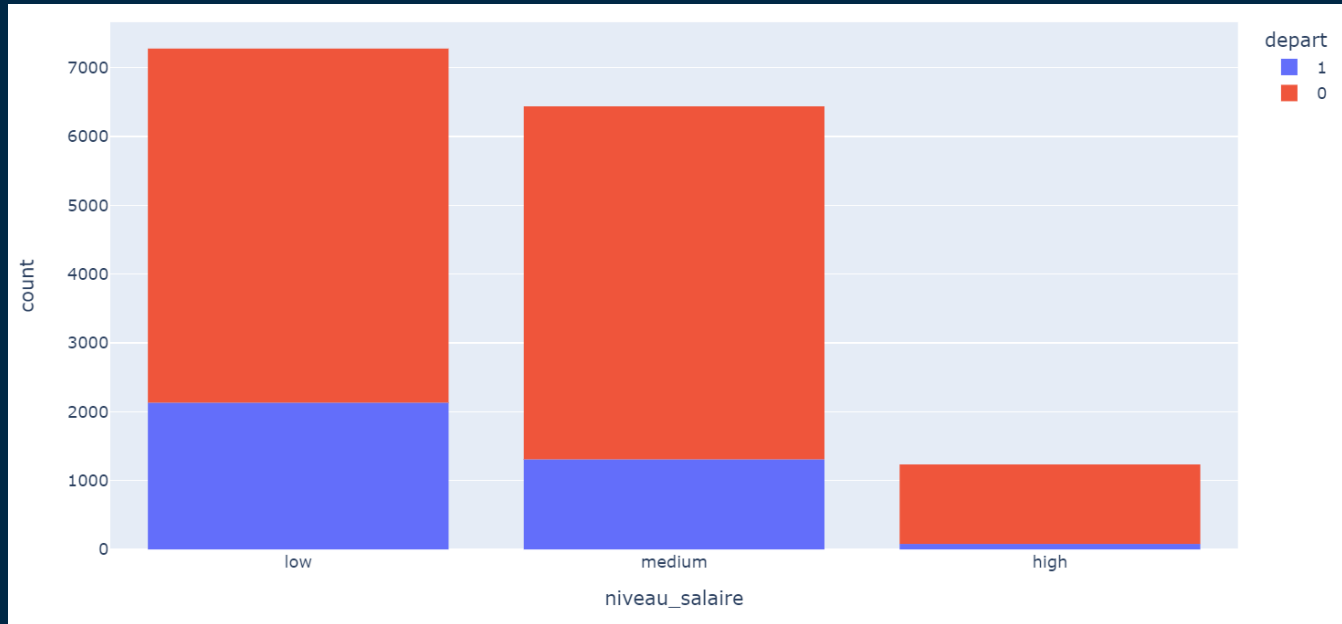
01 JEU DE DONNEES

Un peu de Data Viz



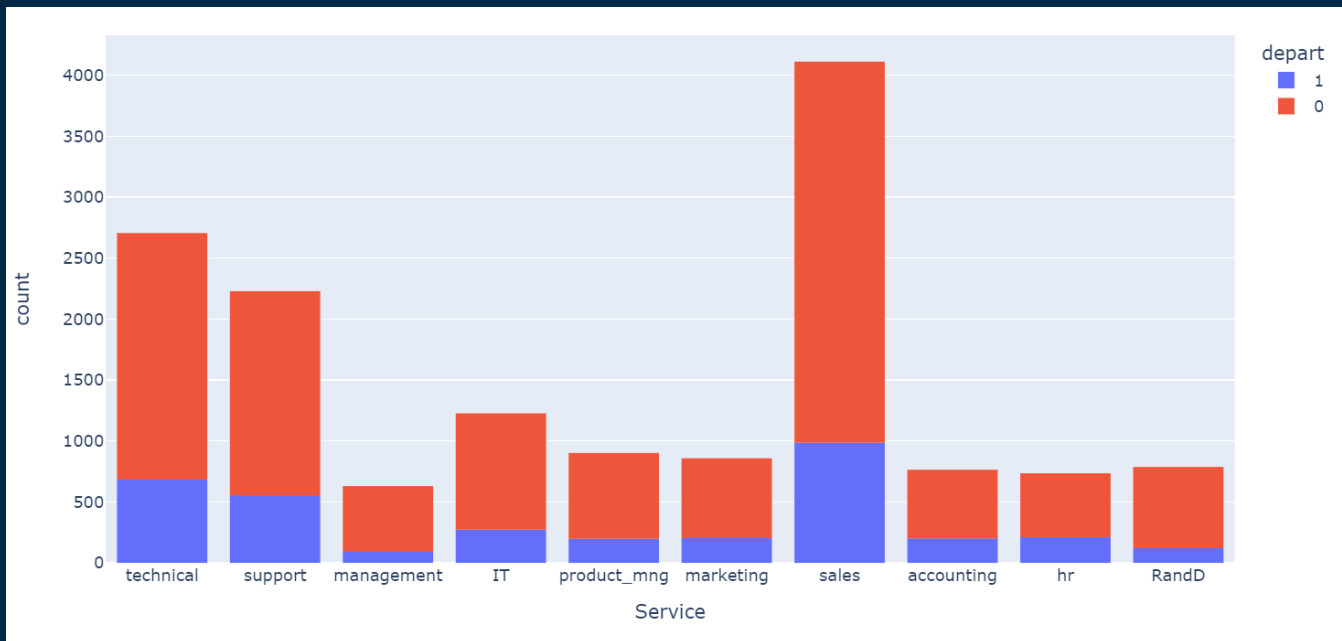
01 JEU DE DONNEES

Répartition des départs en fonction des niveaux de salaire



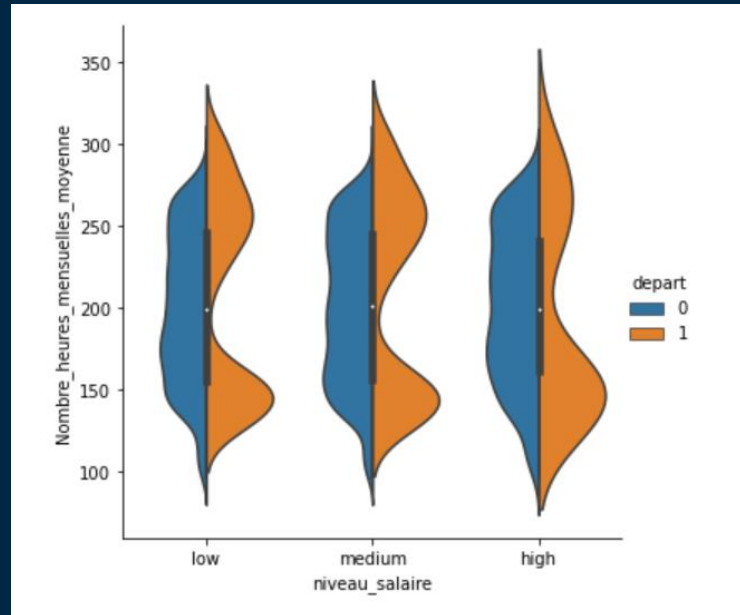
01 JEU DE DONNEES

Répartition des départs en fonction des services



01 JEU DE DONNEES

Répartition des départs en fonction des heures mensuelles travaillées et du niveau de salaire associé



02 METHODOLOGIE



DataBaseTrain



DataBaseTest



Projet_turnover.ipynb



README.md



.ipynb_checkpoints



DataBaseTest



DataBaseTrain



Projet_turnover.ipynb



README.md



decision_tree.png



test.dot

Architecture du projet sous GITHUB



02 METHODOLOGIE

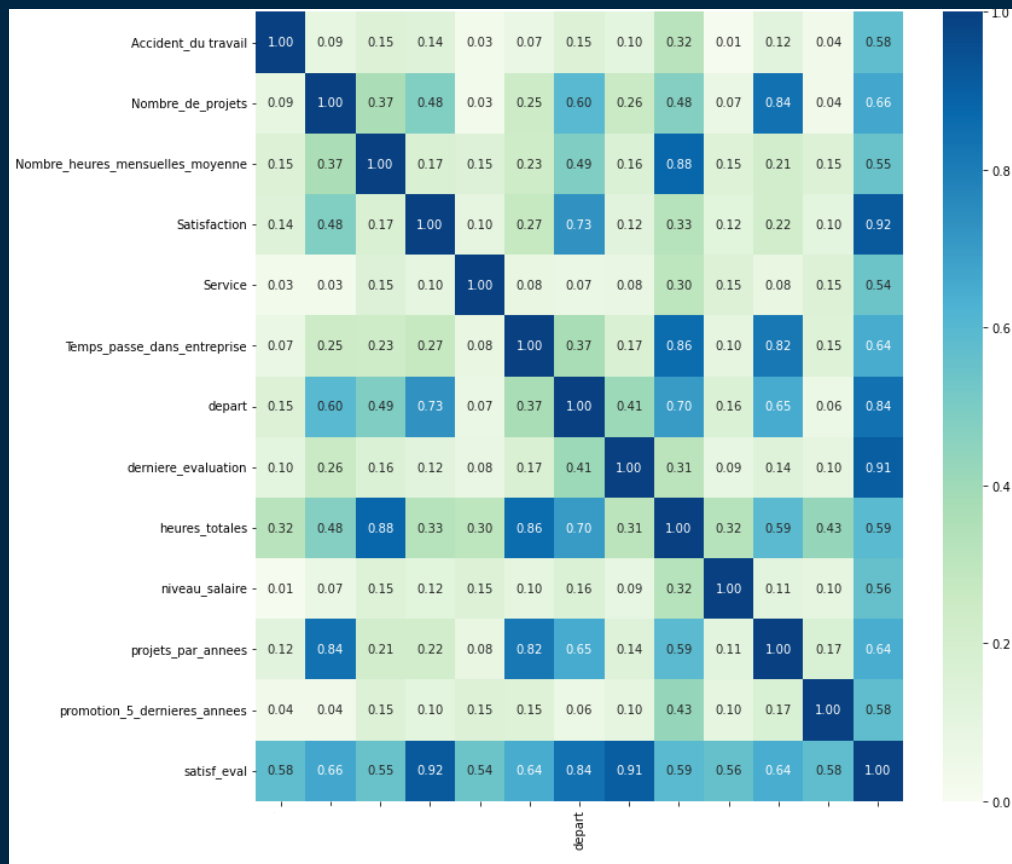
- ETAPE 1 – Data Viz
- ETAPE 2 – Etude de corrélation
- ETAPE 3 – 4 modèles : Random Forest / Logistic Regression / KNN / SVM
- ETAPE 4 – Constat
- ETAPE 5 – Optimisation / Approfondissement



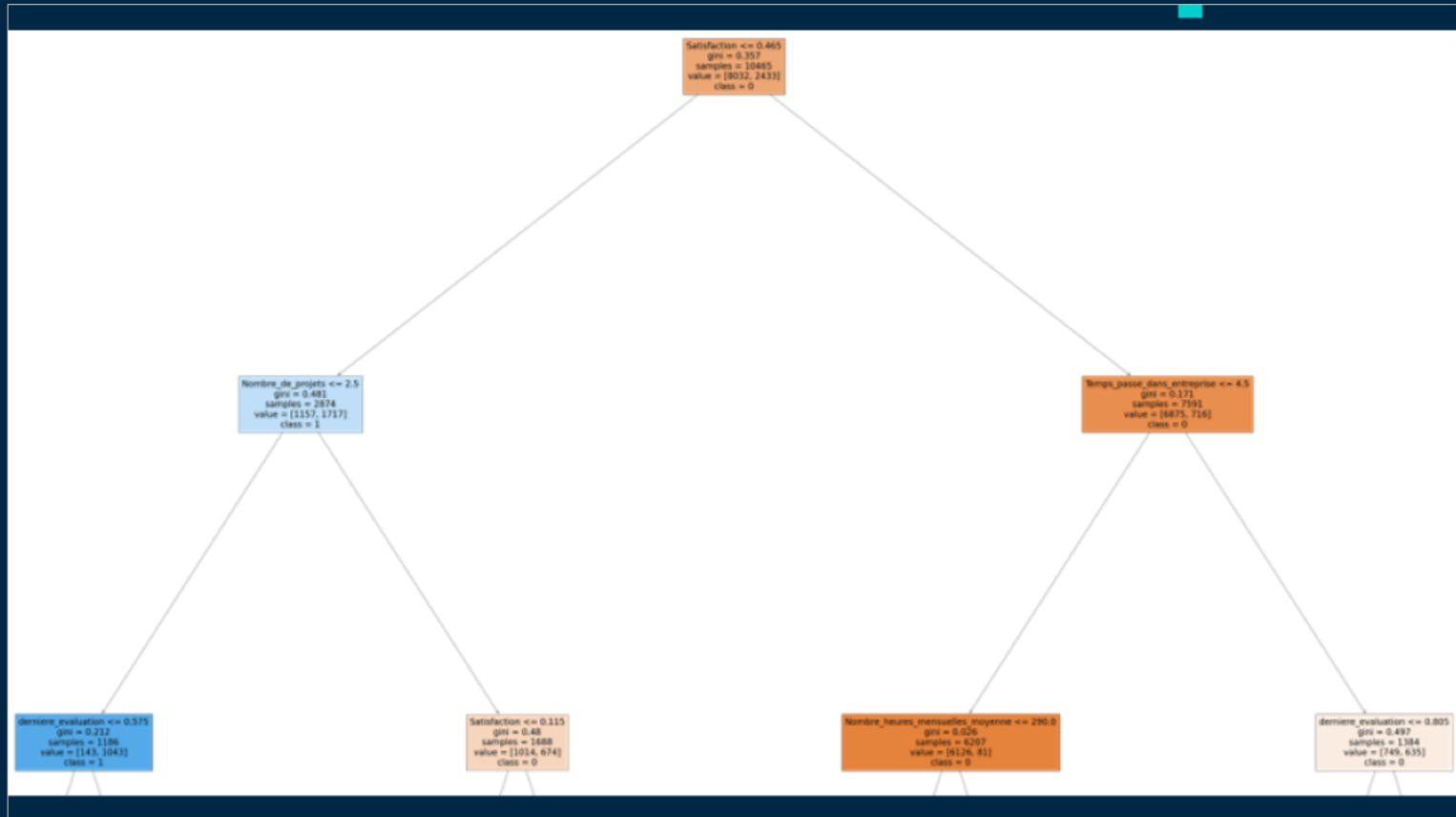
03 MODELES

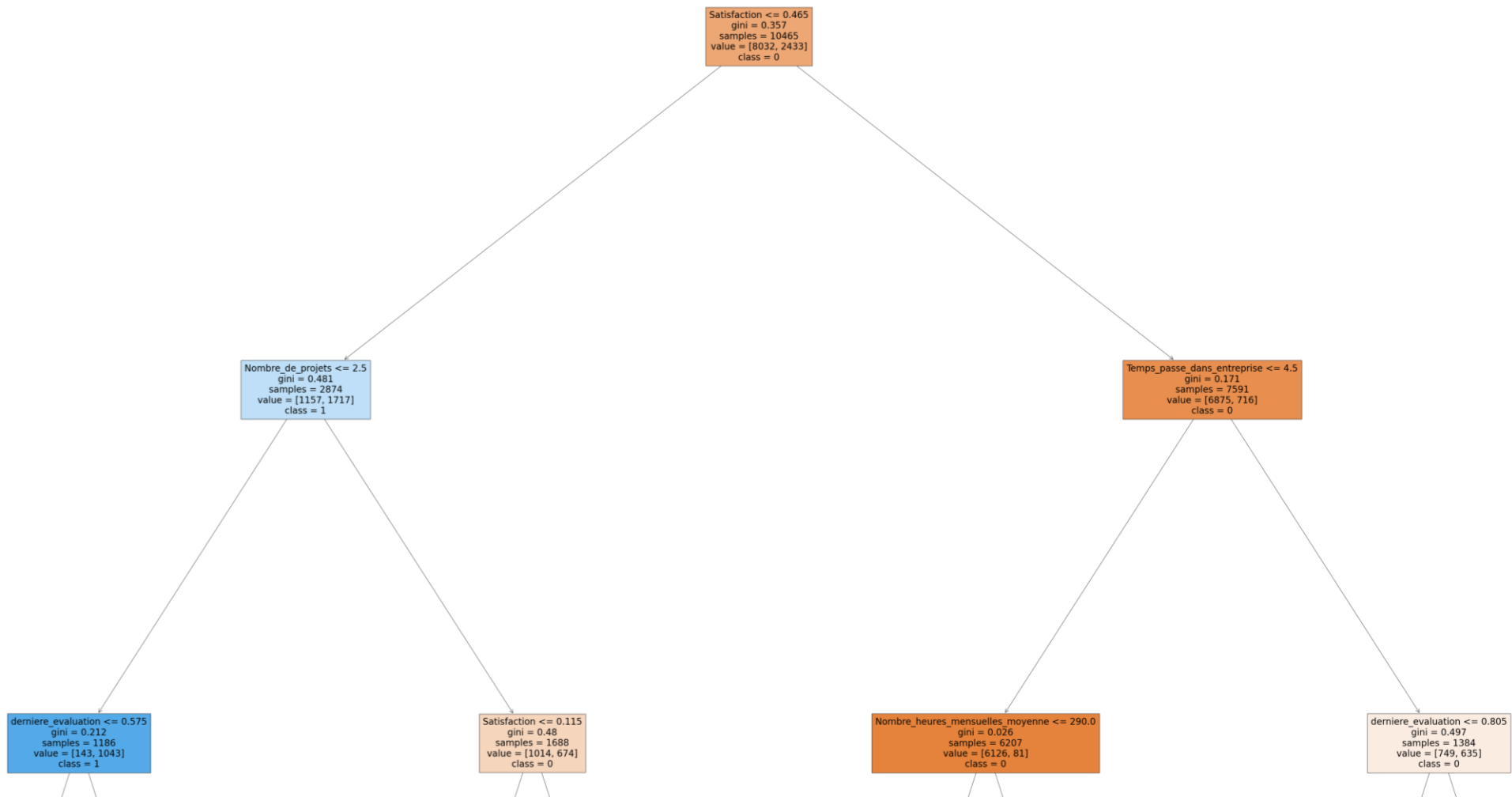


03 MODELES



03 MODELES





03 MODELES

■ Random Forest

- Plusieurs arbres de décisions / Chacun produit une estimation

■ Logistic Regression

- Permet de prédire qu'un élément arrive (0 ou 1)

■ KNN

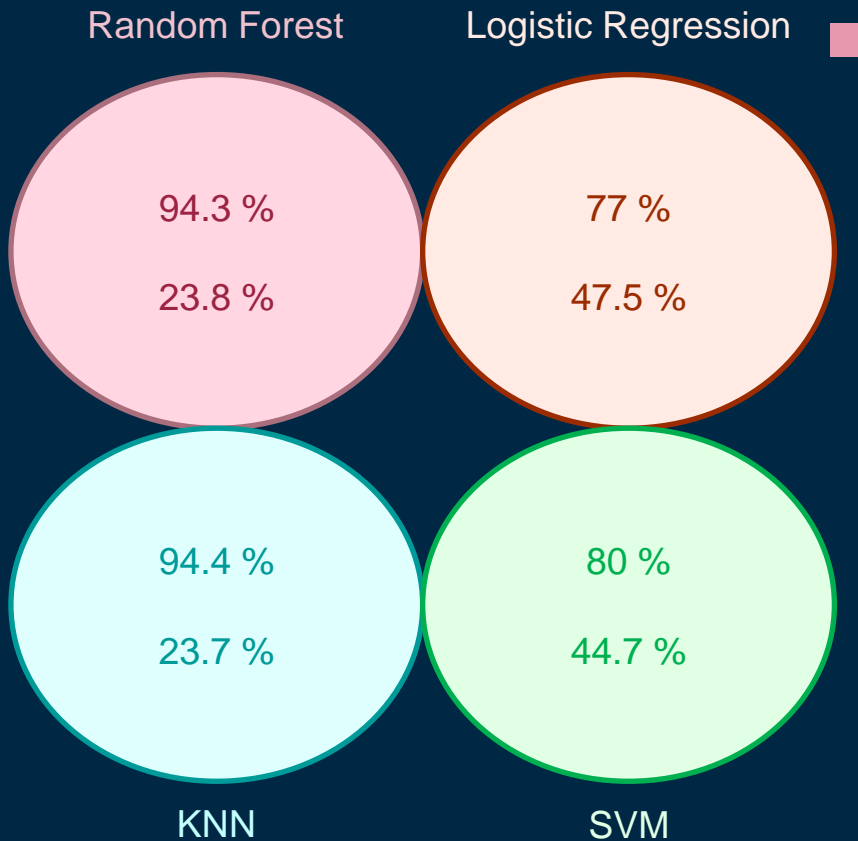
- Classifie des points dans des classes connues en fonction de leur distance / + pour résoudre classification

■ SVM

- Apprentissage automatique / Résoudre des problèmes de classification

04 RESULTATS

Accuracy
RMSE

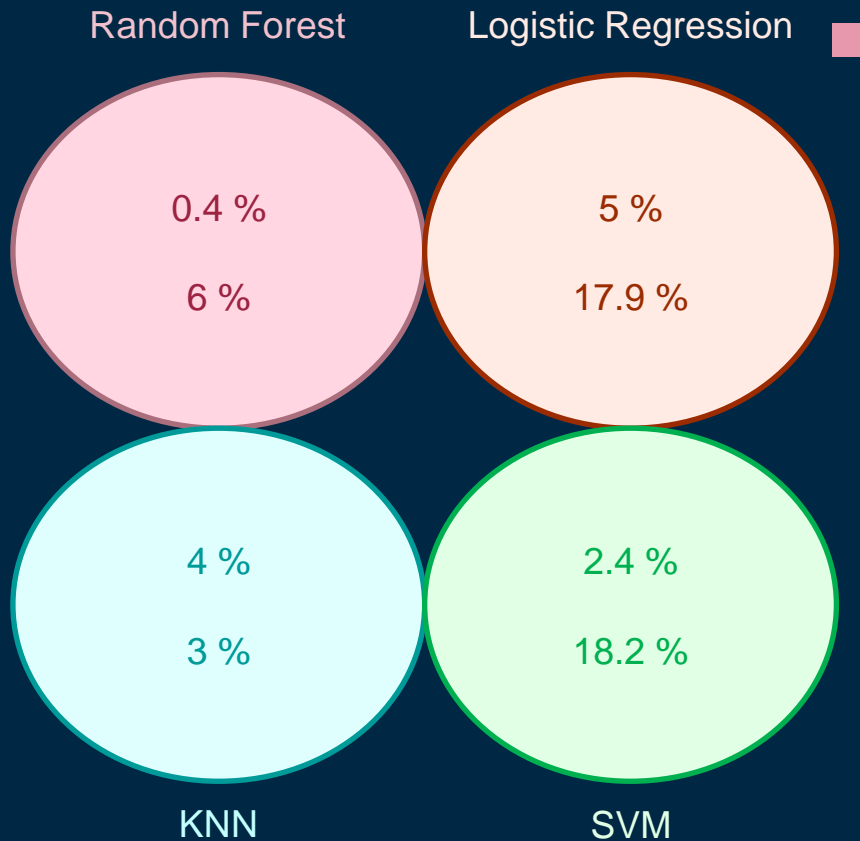


04 RESULTATS

Satisfaction	derniere_evaluation	Nombre_de_clients	Nombre_heures	Temps_passes	Accident_du_travail	promotion_5ans	Service_client	niveau_salaire	random_forest	logistic_regression	knn	svm	expected
0.45	0.54	2	135	3	0	0	1	1	Left	Stay	Left	Stay	Left
0.11	0.81	6	305	4	0	0	1	1	Left	Left	Left	Left	Left
0.84	0.92	4	234	5	0	0	1	1	Stay	Stay	Left	Stay	Left
0.41	0.55	2	148	3	0	0	1	1	Left	Stay	Left	Stay	Left
0.36	0.56	2	137	3	0	0	1	1	Left	Stay	Left	Stay	Left
0.38	0.54	2	143	3	0	0	1	1	Left	Stay	Left	Stay	Left
0.45	0.47	2	160	3	0	0	1	1	Left	Stay	Left	Stay	Left
0.78	0.99	4	255	6	0	0	1	1	Stay	Stay	Stay	Stay	Left
0.45	0.51	2	160	3	1	1	1	1	Left	Stay	Left	Stay	Left
0.76	0.89	5	262	5	0	0	1	1	Left	Stay	Left	Stay	Left
0.44	0.51	2	156	3	0	0	9	3	Left	Stay	Left	Stay	Left
0.09	0.8	7	283	5	0	0	9	1	Left	Left	Left	Left	Left
0.92	0.87	4	226	6	1	0	9	2	Stay	Stay	Left	Stay	Left
0.74	0.91	4	232	5	0	0	9	2	Stay	Stay	Left	Stay	Left
0.09	0.82	6	249	4	0	0	9	2	Left	Left	Left	Left	Left
0.89	0.95	4	275	5	0	0	9	2	Stay	Stay	Left	Stay	Left
0.1	0.86	6	278	4	0	0	9	3	Left	Left	Left	Left	Left
0.81	1	4	253	5	0	0	9	1	Stay	Stay	Left	Stay	Left
0.11	0.8	6	282	4	0	0	9	2	Left	Left	Left	Left	Left
0.11	0.84	7	264	4	0	0	9	2	Left	Left	Left	Left	Left

04 RESULTATS

Faux départ
Faux non départ

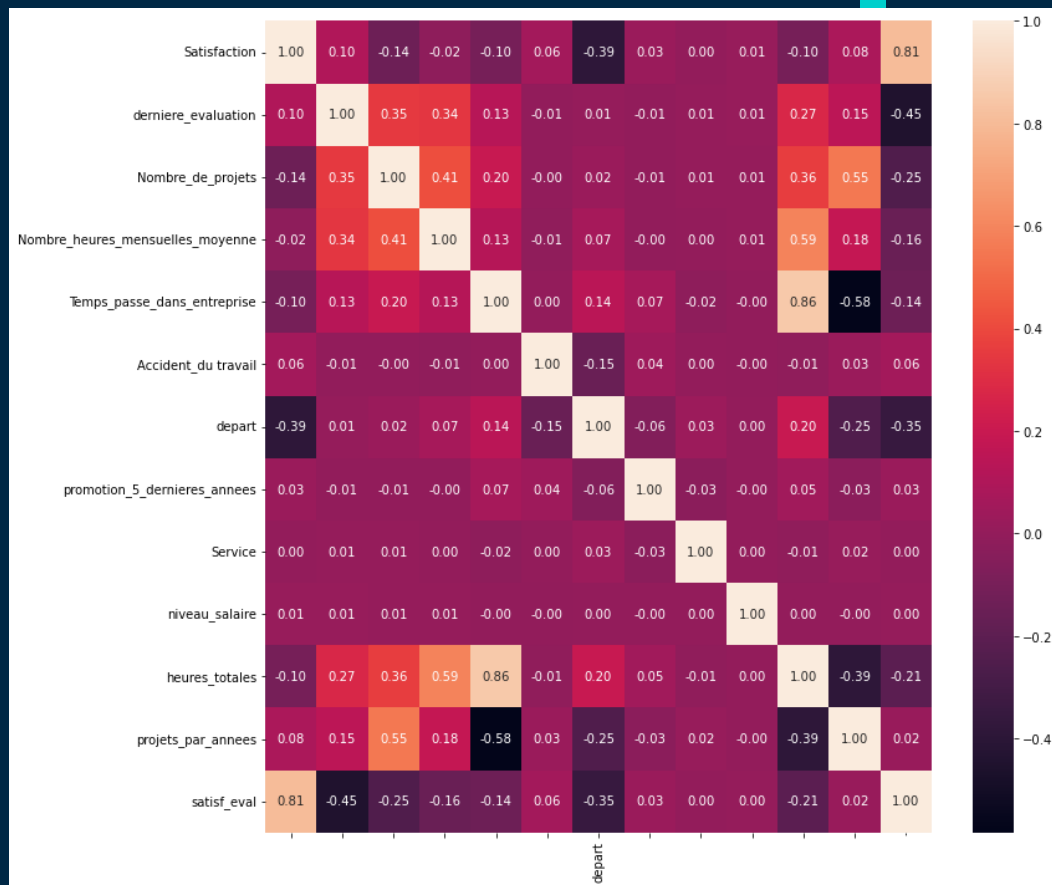


05 APPROFONDIR

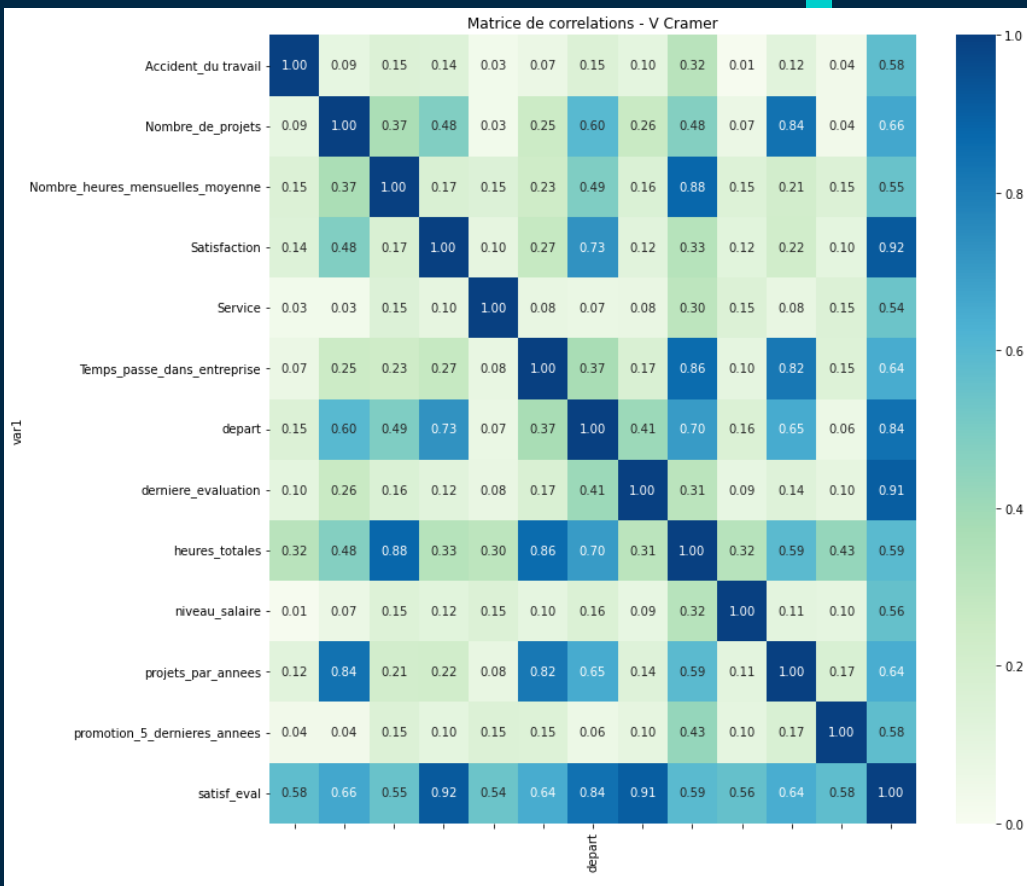
Création de 3 nouveaux éléments :

- `'heures_totales'` = `'Nombre_heures_mensuelles_moyenne'` x 12 x `'Temps_passe_dans_entreprise'`
- `'projets_par_annees'` = `'Nombre de projets'` / `'Temps_passe_dans_entreprise'`
- `'satisf_eval'` = `'Satisfaction'` / `'derniere_evaluation'`

05 APPROFONDIR

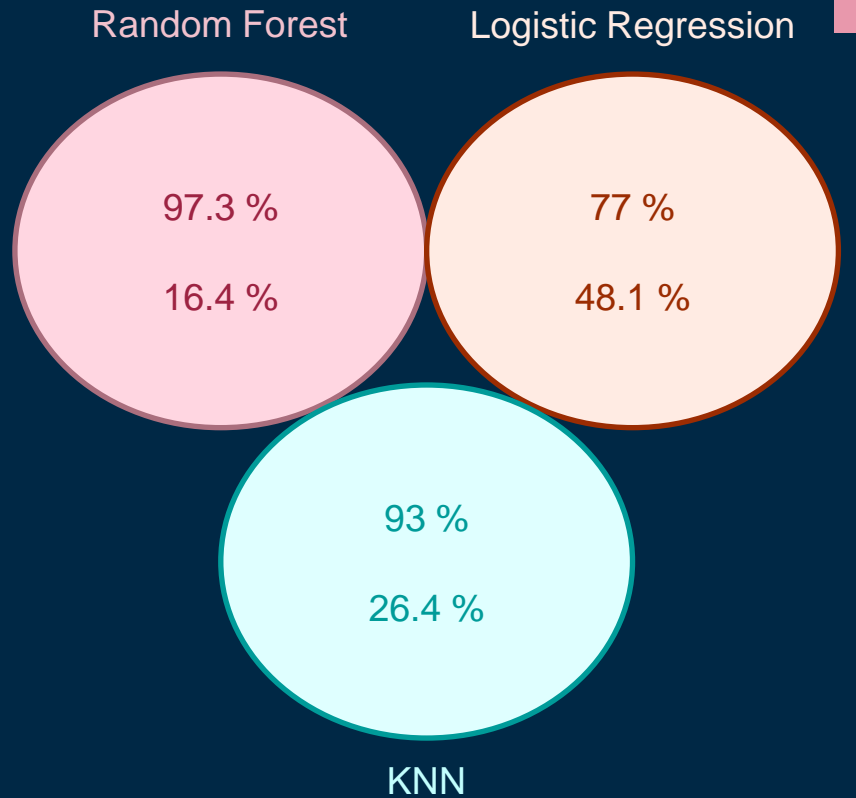


05 APPROFONDIR



05 APPROFONDIR

Accuracy
RMSE

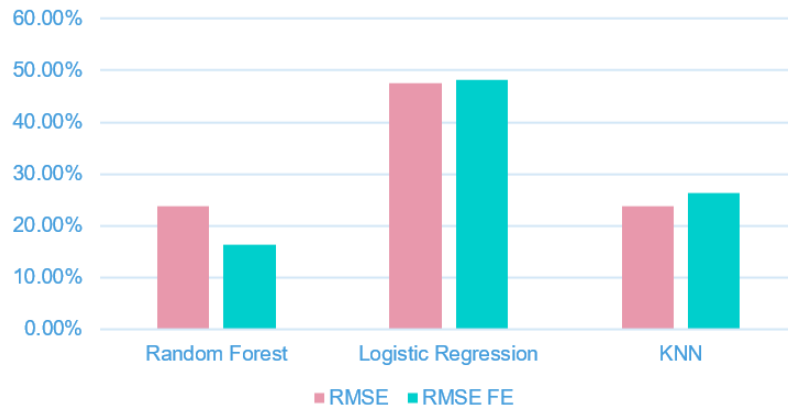


05 APPROFONDIR

Comparaison des Accuracy



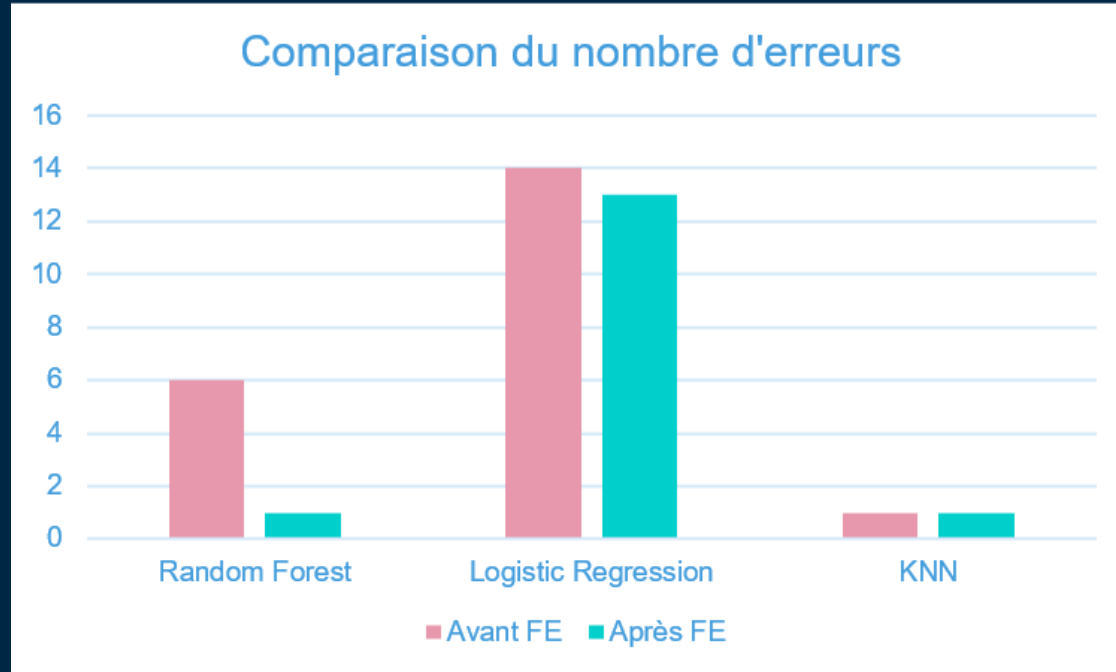
Comparaison des RMSE




05 APPROFONDIR

Satisfaction	Accident_du	heures_total	projets_par	satisf_eval	random fore	logistic regre	knn	expected
0.45	0	4860	0.66	0.83	Left	Stay	Left	Left
0.11	0	14640	1.5	0.13	Left	Left	Left	Left
0.84	0	14040	0.8	0.91	Left	Stay	Left	Left
0.41	0	5328	0.66	0.74	Left	Stay	Left	Left
0.36	0	4932	0.66	0.64	Left	Left	Left	Left
0.38	0	5148	0.66	0.7	Left	Left	Left	Left
0.45	0	5760	0.66	0.95	Left	Stay	Left	Left
0.78	0	18360	0.66	0.78	Left	Stay	Stay	Left
0.45	1	5760	0.66	0.88	Left	Stay	Left	Left
0.76	0	15720	1	0.85	Left	Stay	Left	Left
0.44	0	5616	0.66	0.86	Left	Stay	Left	Left
0.09	0	16980	1.4	0.11	Left	Left	Left	Left
0.92	1	16272	0.66	1.05	Stay	Stay	Left	Left
0.74	0	13920	0.8	0.81	Left	Stay	Left	Left
0.09	0	11952	1.5	0.11	Left	Left	Left	Left
0.89	0	16500	0.8	0.93	Left	Stay	Left	Left
0.1	0	13344	1.5	0.11	Left	Left	Left	Left
0.81	0	15180	0.8	0.81	Left	Stay	Left	Left
0.11	0	13536	1.5	0.13	Left	Left	Left	Left
0.11	0	12672	1.75	0.13	Left	Stay	Left	Left

05 APPROFONDIR



06 CONCLUSION

- Le meilleur est KNN. 
- Le feature engineering n'est pas vraiment nécessaire dans notre cas d'après notre étude.
- Utiliser d'autres modèles pour essayer.

MERCI DE VOTRE ATTENTION

QUESNEL Ninon – BERTON Léonie – REVERSAC Paul – PERICHON Nicolas