

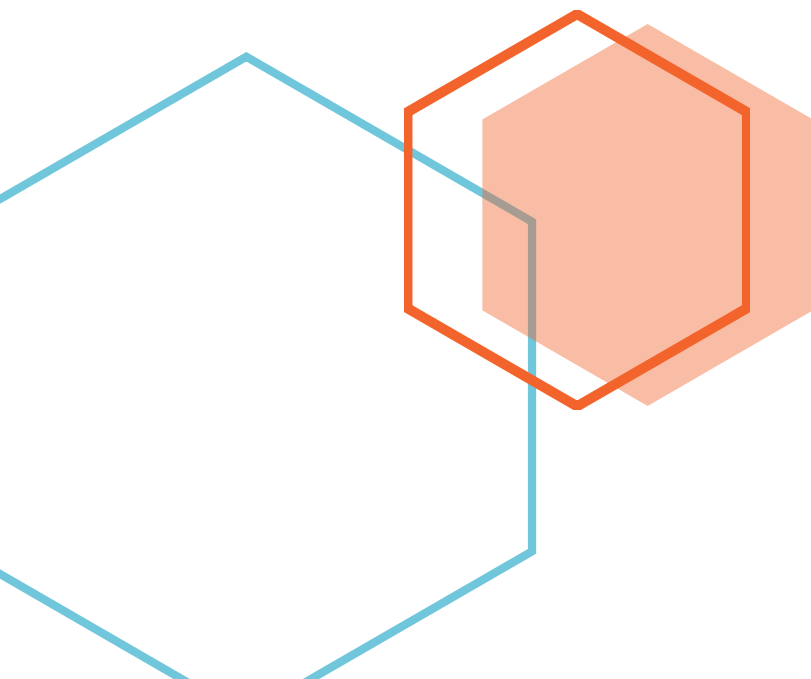


Prévision de nombre des étudiants présents dans les restaurants universitaires

Méthode de classification binaire

Résultats obtenus

Le modèle a été entraîné sur un ensemble de données de 2ans sur une population de 200 étudiants. Le test a été effectué sur une période d'un an avec 3 algorithmes de prédiction différent qui sont La régression logistique, les forets aléatoires et K plus proches voisins KNN



Les forets aléatoire :

Score obtenu : 96.86 %

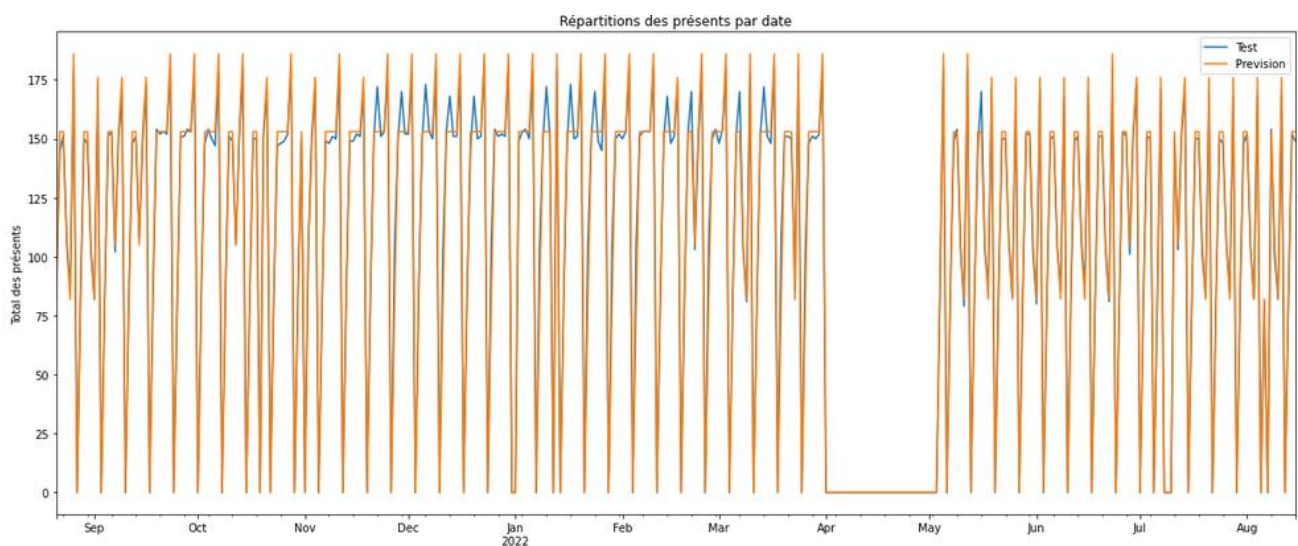
Importance de variables :

importance	
Rev_par	0.292875
Plat	0.274655
Event	0.212606
Heberg	0.122957
Bourse	0.058999
meteo	0.037908

Matrice de confusion :

	blanc_predit	rouge_predit
blanc_données	33185	1166
rouge_données	1094	36554

Aperçu sur les résultats prédits (jeu de test)



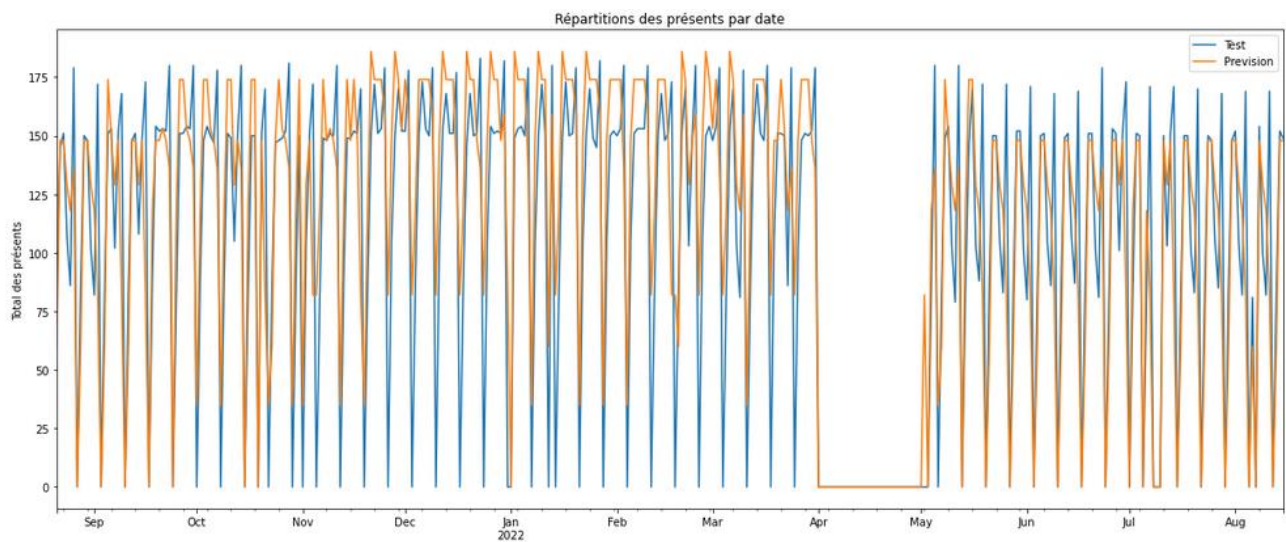
La régression Logistique:

Score obtenu : 84.61%

Matrice de confusion :

	blanc_predit	rouge_predit
blanc_données	27612	6739
rouge_données	4340	33308

Aperçu sur les résultats prédits (jeu de test)





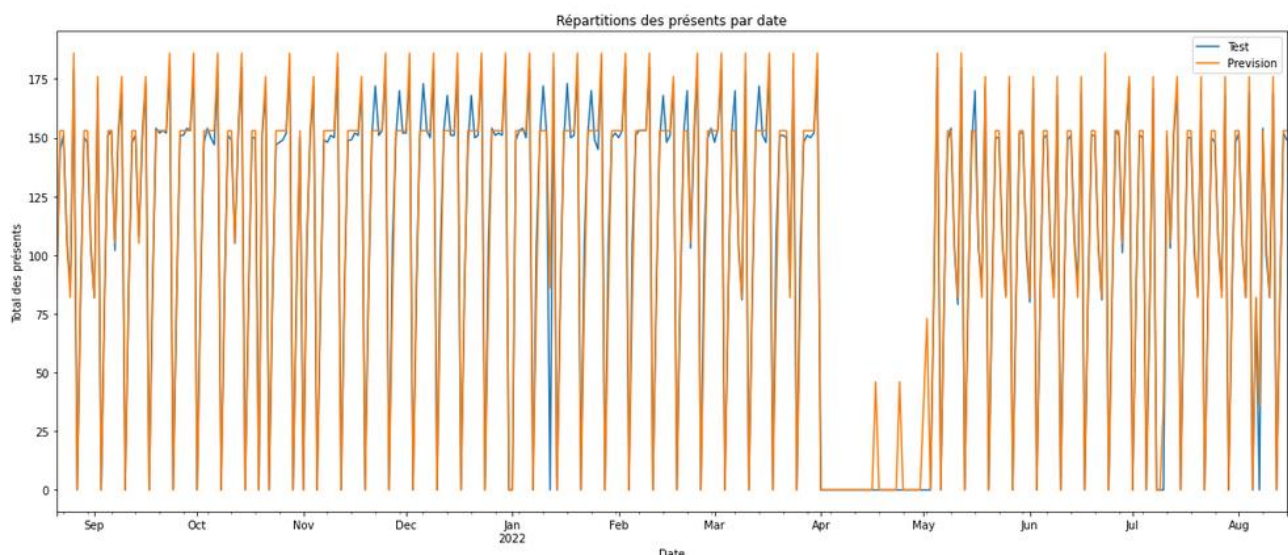
K plus proches voisins KNN:

Score obtenu : 96.36%

Matrice de confusion :

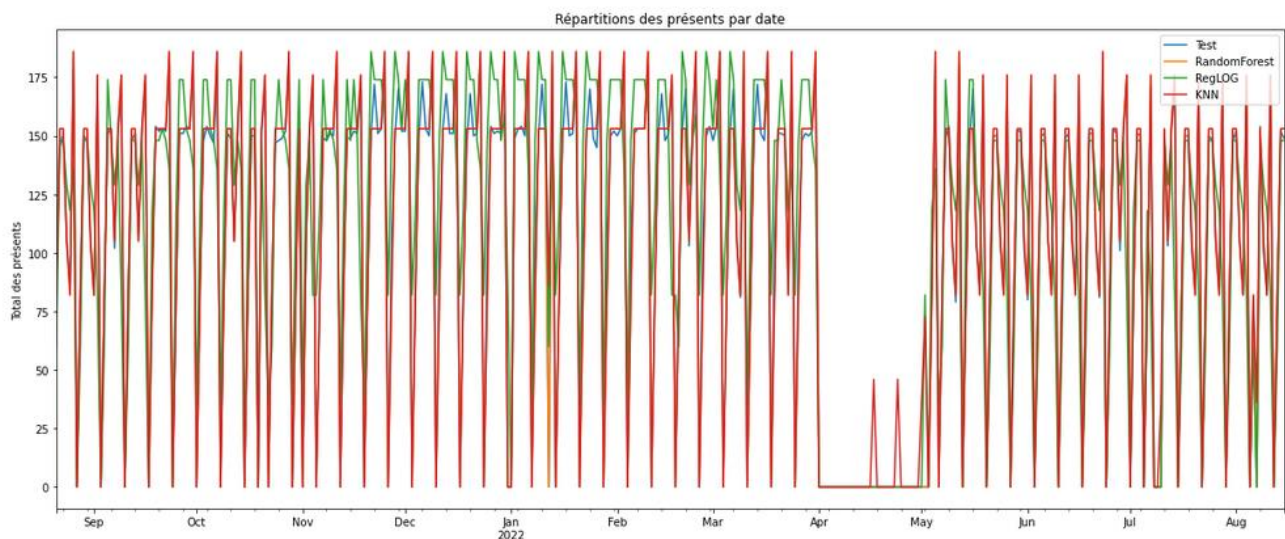
	blanc_predit	rouge_predit
blanc_données	32826	1525
rouge_données	1094	36554

Aperçu sur les résultats prédits (jeu de test)



Résultats finaux :

	Date	Test	RandomForest	RegLOG	KNN
0	2021-08-21	86	82	60	82
1	2021-08-22	145	153	148	153
2	2021-08-23	151	153	148	153
3	2021-08-24	107	105	129	105
4	2021-08-25	86	82	118	82
...
355	2022-08-11	169	176	82	176
356	2022-08-12	0	0	0	0
357	2022-08-13	83	82	60	82
358	2022-08-14	152	153	148	153
359	2022-08-15	149	153	148	153



Conclusion :

- Les deux algorithmes Forêts aléatoire et KNN donnent de bons résultats par rapport à la régression logistique.
- La régression logistique ne possède pas des hyperparamètres à ajuster contrairement aux forêts aléatoires qui dépendent fortement de nombres des arbres et le Knn au nombre des voisins.