

22 . K-Means (Visualisation du modèle) :

On constate bien nos 2 groupes sur nos 2 premières composantes !

-Les centroïdes créés par le K-Means sont proches de ceux des données de base

-Les groupes sont très ressemblants à ceux des données de base

-La zone "d'incertitude" n'existe plus ! On voit une limite entre les 2 groupes !

23 . K-Means (Evaluation du modèle) :

La matrice de confusion nous permet d'évaluer les résultats du K-Means par rapport à nos données de base !

- 163 billets ont été considérés comme vrais alors qu'ils étaient faux ! (32.93% des données)
- 320 billets ont été considérés comme faux alors qu'ils étaient vrais ! (64.65% des données)
- Le taux de données mal prédites est de 97,60 %, soit seulement 483 erreurs sur 495 échantillons.
- 4 billets ont été correctement identifiés comme vrais (0.81% des données)
- 8 billets ont été correctement identifiés comme faux (1.62% des données)
- 97,60% de prédictions incorrectes ! = somme des erreurs 32.93% + 64.65% = 97,60%

Accuracy : Cet indicateur nous indique le pourcentage de prédictions correctes que notre modèle a réalisées sur l'ensemble de données de test. Plus la précision est élevée, meilleur est notre modèle.

inAccuracy : 97,60% de prédictions incorrectes ! et une Accuracy de 2.40%

K-Means, étant un algorithme non supervisé, il ne peut pas rivaliser avec des méthodes supervisées comme la régression logistique pour une tâche de classification

24 . Régression Logistique :

La matrice de confusion nous permet d'évaluer les résultat de la régression logistique par rapport à nos données de base !

-4 Billets ont été considérés comme faux alors qu'ils étaient vrais (0.81% des données, "faux négatifs").

- Le taux de données mal prédites est de 0.81 %, soit seulement 4 erreurs sur 495 échantillons.

-332 billets ont été correctement identifiés comme vrais (67.07% des données)

-159 billets ont été correctement identifiés comme faux (32.12% des données)

Accuracy : Cet indicateur nous indique le pourcentage de prédictions correctes que notre modèle a réalisé sur l'ensemble de données de test. Plus la précision est élevée, meilleur est notre modèle.

- **Accuracy : 99.12 % de prédictions correctes !**
- **Inaccuracy : 0.81 % de prédictions incorrectes!**

25 . Lecture de la Courbe ROC

La courbe ROC est une représentation graphique de la performance de notre modèle en fonction de différents seuils de classification.

Elle trace la sensibilité en fonction de la spécificité pour chaque seuil de classification possible.

L'aire sous la courbe ROC mesure la qualité globale de notre modèle. Plus l'AUC est proche de 1, meilleur est notre modèle.

On tombe sur un modèle quasiment parfait avec une aire sous la courbe approximée à 1