### ****1. Analyse exploratoire & corrélation****

**Q : Pourquoi faire une matrice de corrélation ?**  
**R :** Pour détecter les variables fortement liées entre elles. Cela aide à identifier la redondance (ex. height\_left et height\_right) et à comprendre les relations structurelles dans les données.

**Q : Problème si les variables sont corrélées ?**  
**R :**

* Pour les modèles linéaires (ex. régression logistique) → multicolinéarité → coefficients instables.
* Pour Random Forest → pas un problème majeur, mais certaines variables peuvent être jugées moins importantes car déjà représentées par d'autres.

### ****2. Modèles testés****

**Q : Quels modèles as-tu testés ?**  
**R :**

**Régression Logistique** : rapide, interprétable, mais sensible aux variables corrélées et à l’échelle.

**K-means** : non supervisé, moins performant pour prédire is\_genuine.

**KNN** : simple, basé sur la proximité, sensible à la normalisation et au choix du k.

**Random Forest** : robuste, performant, interprétable via importance des features.

### ****3. Pourquoi choisir Random Forest****

**Q : Pourquoi Random Forest est le meilleur choix ici ?**  
**R :**

* Gère les variables corrélées.
* Pas besoin de scaling.
* Robuste au bruit et au sur-apprentissage grâce au bagging.
* Offre une mesure d’importance des variables → utile pour expliquer le modèle.

### ****4. Interprétation des résultats****

**Q : Comment interpréter l’importance des variables dans Random Forest ?**  
**R :** Plus la valeur d’importance est élevée, plus la variable contribue à séparer correctement les classes (is\_genuine). Exemple : diagonal peut avoir plus de poids que margin\_low.

**Q : Si deux variables sont corrélées, comment le modèle choisit ?**  
**R :** Il répartit l’importance entre elles, mais privilégie celle qui apporte le plus d’information unique lors des splits d’arbres.

### ****5. Améliorations possibles****

**Q : Comment améliorer les performances ?**  
**R :**

* Ajuster hyperparamètres (n\_estimators, max\_depth, min\_samples\_split).
* Faire une validation croisée pour évaluer la robustesse.
* Éventuellement supprimer ou combiner certaines variables fortement corrélées.