

Détection automatique des faux billets

Projet de classification supervisée avec application
fonctionnelle

Théo ZWAHLEN

Contexte et objectif du projet

Données fournies :

- 1500 billets au total :
1000 vrais, 500 faux

Chaque billet est décrit par 6 mesures physiques :





- diagonal
- height_left, height_right
- margin_low, margin_up
- length

Le projet consiste à :

- Tester différents algorithmes de classification
- Créer une application fonctionnelle pour prédire de nouveaux billets







Analyse exploratoire des données

- Analyse de la distribution des variables géométriques
 - Aucune anomalie majeure détectée
 - **37 valeurs manquantes** dans `margin_low`
 - Comparaison des variables selon la classe `is_genuine` :
 - Faux billets = **`margin_low`** plus élevée, **`length`** plus courte
 - Autres variables : différences plus modérées mais significatives
 - Certaines variables sont discriminantes → utiles pour les modèles
- 
- 
- 
- 







Traitement des valeurs manquantes et standardisation

- 
- 
- Utilisation d'une **régression linéaire** pour estimer les valeurs manquantes
 - $R^2 \approx 0.48$ → le modèle explique 48 % de la variation
 - Résidus pas parfaitement normaux
 - Homoscédasticité non respectée
-
- Objectif : **imputation**, pas interprétation statistique
 - **Standardisation** appliquée ensuite
- 
- 




Préparation des données pour l'apprentissage supervisé

- Variable cible : is_genuine (booléen : True = vrai billet, False = faux)
 - Variables explicatives : les 6 dimensions géométriques
 - **Séparation du jeu de données :**
 - 80 % pour l'entraînement
 - 20 % pour le test
 - Standardisation : mise à l'échelle des variables
 - Objectif : entraîner plusieurs modèles et comparer leur performance sur la même base
- 
- 
- 
- 



Modèles de classification testés

Modèle	Type	Description rapide
Régression logistique	Linéaire supervisé	Classifieur probabiliste, adapté aux données séparables
KNN	Non paramétrique supervisé	Classe selon les échantillons les plus proches
Random Forest	Arbre supervisé	Agrégation de plusieurs arbres de décision aléatoires
K-Means	Non supervisé	Regroupe en k culsters selon distance aux centroïdes



Résultats des modèles - Évaluation comparée

Modèle	Accuracy	F1-score	Observations clés
Régression logistique	0.99	0.99	Excellente performance, très bon équilibre précision/rappel
KNN	0.98	0.98	Bon modèle mais légèrement moins performant, 5 faux billets non détectés
Random Forest	0.99	0.99	Très robuste, aussi performant que la régression logistique
K-Means	0.99	0.99	Très bonne séparation malgré l'absence d'étiquettes à l'entraînement



Application fonctionnelle Prédiction en production

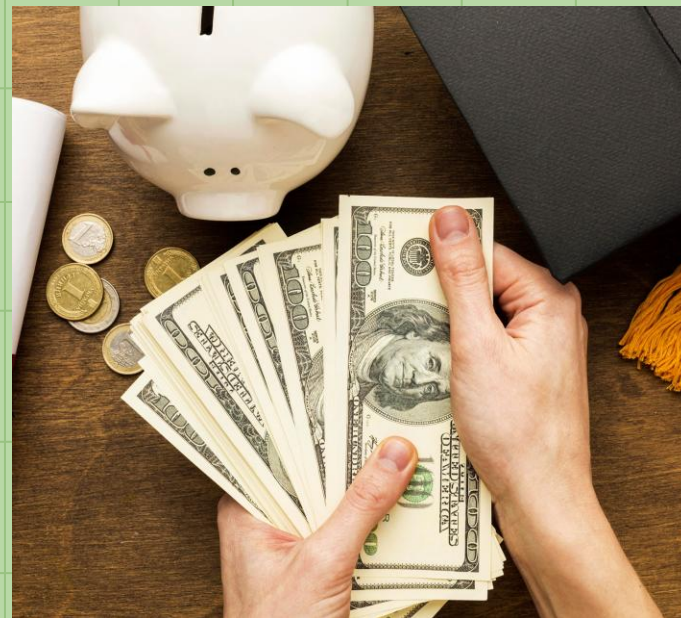
- Application développée dans un notebook dédié
- Chargement du modèle final (Random Forest) et du scaler
- Prédiction à partir d'un fichier CSV contenant les dimensions des billets
- Affichage de la prédiction et de la probabilité : vrai ou faux billet

Objectif atteint : un outil simple pour les équipes de l'ONCFM



Conclusion et recommandations

- Plusieurs modèles testés : **régression logistique, KNN, Random Forest, K-means**
- **Random Forest retenu** : meilleur compromis entre performance et robustesse
- Analyse complète des dimensions des billets
→ signal exploitable
- Application fonctionnelle prête à l'emploi (fichier CSV en entrée)
- Objectif du projet atteint : assistance rapide à la détection de faux billets



The background of the slide features a light green grid pattern. Scattered across this grid are numerous green dollar bills, depicted in a stylized, cartoonish manner. Each bill has a white outline and a large green dollar sign in the center. The bills are oriented in various directions, some appearing to be falling or floating.

Merci

Avez-vous des questions ?