



Détectez des faux billets

[MISSION](#)[COURS](#)[ÉVALUATION](#)

80 heures

Mis à jour le mardi 2 avril 2019

Prérequis

Pour ce projet, il sera nécessaire de savoir manipuler des données en **R** ou en **Python**. Également, vous devrez effectuer une analyse de **statistique descriptive**, ainsi qu'une **Analyse en Composantes Principales**, une **classification automatique**, et une modélisation de type **régression logistique**.

Mise en situation

Votre société de consulting informatique vous propose une nouvelle mission au ministère de l'intérieur, dans le cadre de la lutte contre la criminalité organisée, à l'[office central pour la répression du faux monnayage](#). Votre mission si vous l'acceptez : créer un algorithme de détection de faux billets.

Vous vous voyez déjà en grand justicier combattant sans relâche la criminalité organisée en pianotant à mains de maîtres votre ordinateur, pour façonner ce fabuleux algorithme qui traquera la moindre fraude et permettra de mettre à jour les réseaux secrets de fossoyeurs ! La classe, non ?

... bon, si on retombait les pieds sur terre? Travailler pour la police judiciaire, c'est bien, mais vous allez devoir faire appel à vos connaissances en statistiques, alors on y va !

Les données

La PJ vous transmet un jeu de données contenant les caractéristiques géométriques de billets de banque. Pour chacun d'eux, nous connaissons :

- La longueur du billet (en mm)
- La hauteur du billet (mesurée sur le côté gauche, en mm)
- La hauteur du billet (mesurée sur le côté droit, en mm)
- La marge entre le bord supérieur du billet et l'image de celui-ci (en mm)
- La marge entre le bord inférieur du billet et l'image de celui-ci (en mm)
- La diagonale du billet (en mm)

[Cliquez pour télécharger billets.csv](#)

Votre mission

Mission 0

Afin d'introduire votre analyse, effectuez une brève description des données (analyses univariées et bivariées).

Mission 1

Vous réaliserez une Analyse en Composantes Principales de l'échantillon, en suivant toutes ces étapes :

- Analyse de l'éboulis des valeurs propres
- Représentation des variables par le cercle des corrélations
- Représentation des individus par les plans factoriels
- Analyser la qualité de représentation et la contribution des individus.

Pour chacune de ces étapes, commentez les résultats obtenus. La variable donnant la nature Vrai/Faux du billet sera utilisée comme variable illustrative.

Mission 2

Appliquer un algorithme de classification, puis analysez le résultat obtenu.

Visualisez la partition obtenue dans le premier plan factoriel de l'ACP, puis analysez-la.

Mission 3

Modélisez les données à l'aide d'une régression logistique. Grâce à celle-ci, vous créerez un programme capable d'effectuer une prédiction sur un billet, c'est-à-dire de déterminer s'il s'agit d'un vrai ou d'un faux billet. Pour chaque billet, votre algorithme de classification devra donner la probabilité que le billet soit vrai. Si cette probabilité est supérieure ou égale à 0.5, le billet sera considéré comme vrai. Dans le cas contraire, il sera considéré comme faux.



Il vous sera demandé d'utiliser ce programme lors de votre soutenance !

Livrables attendus

Le livrable attendu est un fichier .zip contenant :

- Le code Python ou R de l'analyse complète (un ou plusieurs fichiers)
- Les graphiques générés par le code, en format image

Modalités de la soutenance

Lors de la soutenance, l'examineur vous procurera un fichier CSV contenant 5 caractéristiques de billets. Votre programme devra déterminer si chacun de ces billets est vrai ou faux, en affichant la probabilité associée. Ce fichier sera de la même forme que celui-ci : [cliquez pour télécharger exemple.csv](#)

Dans ce fichier d'exemple, chaque billet possède un identifiant (dans colonne "id"). Vérifiez que votre programme affiche bien cet identifiant pour chaque prédiction effectuée.

Assurez-vous que votre programme puisse bien lire ce fichier et effectuer la prédiction pour chacun des billets !

La soutenance se déroulera comme suit :

5 min	Présentation du jeu de données & brèves analyse bivariées et univariées (mission 0)
10 min	ACP : détail de l'analyse, avec graphiques à l'appui (mission 1)
5 min	Analyse de la classification (mission 2)
5 min	Détail de la modélisation (mission 3)
5 min	Test du programme avec le jeu de test donné par le mentor-examineur
5 à 10 min	Séance de questions-réponses éventuelle

Ressources complémentaires

<https://www.youtube.com/watch?v=zAULhNrnuL4>

- Introduction en anglais à la régression logistique.

<https://openclassrooms.com/courses/explorez-vos-donnees-avec-des-algorithmes-non-supervises/partitionnez-vos-donnees-avec-dbscan>

- Cours OpenClassrooms sur l'analyse non supervisée (attention, niveau avancé !)

Compétences évaluées



Interpréter une ACP



Modéliser grâce à la régression logistique



Réaliser une ACP



Utiliser un algorithme de clustering de type Kmeans

OpenClassrooms

- L'entreprise
- Alternance
- Forum
- Blog
- Nous rejoindre

Entreprises

- Business

En plus

- Devenez mentor
- Aide et FAQ
- Conditions Générales d'Utilisation
- Politique de Protection des Données Personnelles
- Nous contacter

 Français 

