



← Mon parcours

Détectez des faux billets

100%

Mission

Cours

Ressources

Évaluation

80 heures

Mis à jour le lundi 20 avril 2020

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**.

Scénario

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 maître votre ordinateur, pour façonner ce fabuleux algorithme qui traquera la moindre fraude et permettra de mettre à jour les réseaux secrets de faux-monnayeurs ! 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).

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 de 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

Appliquez 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 créez 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

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.

Pour faciliter votre passage au jury, déposez sur la plateforme, dans un dossier nommé "*P6_nom_prenom*", tous les livrables du projet. Chaque livrable doit être nommé avec le numéro du projet et selon l'ordre dans lequel il apparaît, par exemple "*P6_01_code*", "*P6_02_graphiques*", et ainsi de suite.

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 example.csv](#).

Dans ce fichier d'exemple, chaque billet possède un identifiant (dans la 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 :

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

Compétences évaluées



Réaliser une ACP



Utiliser un algorithme de clustering de type Kmeans



Modéliser grâce à la régression logistique



Interpréter une ACP