

# Gaspard Palay AutoEvaluation PackageRayShader

Gaspard PALAY

1/30/2021

## 1. Critères d'évaluation

1. Comportement du Rmd lors de son exécution
2. Qualité rédactionnelle du dossier
3. Didactisme et pertinence du dossier
4. Qualité et lisibilité du Rmarkdown
5. Qualité des applications permettant d'illustrer le package

## 2. Lien vers le document commenté

En cliquant **ici**, vous trouverez le lien menant au GitHub hébergeant le fruit de mon travail.

## 3. Points positifs de mon travail

Afin d'évaluer mon travail j'ai dû visualiser la totalité des travaux produits par mes camarades de promotion. Ceux-ci ont fournis des travaux de très bonne qualité qui me serviront pour de prochaines utilisations. Il est donc difficile de se comparer à des travaux tous aussi intéressants les uns que les autres et dans l'ensemble très originaux.

J'ai choisi de travailler sur la régression logistique appliquée au machine learning car je n'avais auparavant jamais entraîné un algorithme. De plus, n'étant pas assez doué en statistique, j'ai préféré produire une partie mathématique que j'ai peu détaillé et m'entraîner à l'application de cette régression sur le jeu de données du Titanic. Cela me permettait de compenser mes faiblesses en statistiques. J'ai cependant réussi à comprendre le fonctionnement général du modèle. Même si ce travail n'est pas au niveau mathématique de certains élèves de ma promotion, je considère qu'il est complet et que la partie R lui apporte une réelle valeur ajoutée.

## 3. Points négatifs de mon travail

Comme expliqué précédemment, la partie mathématique n'est pas assez développée en comparaison aux autres travaux de mes camarades, que j'ai pu étudier. Cela m'a donc fait réaliser que j'avais de nombreuses lacunes en statistiques et qu'il faudra obligatoirement que je rattrape ces retards. Je vais donc reprendre des cours de remise à niveau appliquée aux études que je fais.

La partie machine learning de mon travail est biaisé car j'ai nettoyé le jeu de données sur le full au lieu de le faire sur le test. Cela m'a pris beaucoup de temps de paramétrer les chunk afin que le jeu soit nettoyé, je n'ai donc pas eu le temps de corriger cette faute mais ai cependant averti le lecteur que l'accuracy était artificiellement relevée par ma faute.

Les plots générés sur ce travail ne sont pas très visuels et manquent de paramétrage. Cela s'explique par le fait que je maîtrise pas encore assez bien le package ggplot2, comme ceux qui ont produits des travaux

sur celui ci peuvent le maitriser. D'autres data vis auraient pu être générées notamment sur les variables qualitatives. D'autres diagrammes de densités auraient pu être générés ou un seul comprenant plusieurs variables.

## **6. Evaluation du travail suivant les 5 critères**

### **1. Comportement du Rmd lors de son exécution**

L'exécution du code Rmd se réalise parfaitement et sans conditions. Le code RMD appelle bien les libraries nécessaire et les installe aussi si l'utilisateur ne les possède pas au préalable.

### **2. Qualité de la rédaction du dossier**

La rédaction de ce document est de bonne qualité.

### **3. Didactisme et pertinence du dossier**

Le dossier se retrouve facilement sur mon github qui lui même manque de structure (a rajouter un README et autres suivi des mises à jour). Le fichier ne comporte pas de table des matières.

Les descriptions sont bien explicitées et bien illustrées. De nombrueux commentaires se retrouvent sur ma production.

Je transmets une connaissance nécessaire à la reproduction de mon travail.

### **4. Qualité et lisibilité du RMarkdown**

Le RMarkdown est bien écrit. Les plots se génèrent facilement. Il est lisible et les chunk sont bien paramétrés.

### **5. Qualité des applications permettant d'illustrer le package**

Je m'appuie sur différents tutoriels github et youtube que je cite.

Vous retrouvez ce document sur mon **GitHub**.