

Examen R

Chaymae GASMI

Introduction

Dans ce document , nous allons évaluer 5 travaux de mes collègues en R, pour cela, nous allons introduire les 5 critères d'évaluation choisis:

- **Exécution du Rmd**
- **Qualité de la rédaction**
- **Qualité et lisibilité du RMarkdown**
- **Aspect didactique du dossier**
- **Bibliographie**

Les 5 travaux choisis:

- **Package Rpart**
- **Package Lubridate**
- **Package GGplot2**
- **Package Janitor**
- **Package Rminer**
- **Package “Travail personnel”**

Package Rpart

1. Entête

Nous allons explorer dans ce document le package RPART qui a été expliqué par Maxime et Siva . En cliquant ici vous trouverez le lien menant au GitHub: https://github.com/ARSICMrk/ARSIC_PSBx/tree/main/R_Travail_Sup (https://github.com/ARSICMrk/ARSIC_PSBx/tree/main/R_Travail_Sup).

2. Synthèse

Le package RPART (Recursive Partitioning and Regression Tree) permet d'utiliser un processus en deux étapes pour construire des modèles de classification structurelle ou de régression très généraux. Le modèle résultant peut être exprimé sous la forme d'un arbre de décision. Cette méthode est une technique d'apprentissage automatique prédictive puissante et populaire, également appelée **arbre de classification et de régression** (CART).

En effet l'algorithme est basé sur la construction d'un arbre, ce qui rend la méthode facile à expliquer.

3. Extrait

Maxime et siva ont choisi de travailler sur le dataset de **Ptitanic** disponible avec la librairie Rpart. Ce jeu de données concerne le naufrage d'un Titanic avec à son bord des individus classés par catégorie (variables).

Tout d'abord , les auteurs ont créé un dataset d'apprentissage sur lequel l'algorithme s'entraînera, ainsi qu'un dataset de validation pour tester l'algorithme afin d'évaluer sa performance.

```
nb_lignes <- floor((nrow(ptitanic)*0.75))
ptitanic.apprt <- ptitanic[1:nb_lignes, ]
ptitanic.test <- ptitanic[(nb_lignes+1):nrow(ptitanic), ]
```

En effet, 75% du dataset initial définit l'échantillon d'apprentissage,

Puis, ils définissent l'échantillon de test, **ptitanic.test**, soit 25% du dataset initial.

L'étape suivante consiste à construire un arbre de décision en utilisant l'échantillon d'apprentissage.

```
#construction de L'arbre
ptitanic.Arbre <- rpart(survived~.,data= ptitanic.apprt,
                        control=rpart.control(minsplit=5,cp=0))

#affichage de L'arbre
plot(ptitanic.Arbre, uniform=TRUE, branch=0.5, margin=0.1)
text(ptitanic.Arbre,all=FALSE, use.n=TRUE)
```

-Le `survived~.` indique que les auteurs souhaitent prédire la variable *survived* en fonction de toutes les autres.

-Le `minsplit` permet de découper les feuilles qui contiennent au moins 5 observations.

-Le `cp=0` permet de signifier qu'aucune contrainte de découpage n'est exercée

les auteurs recherchent le `cp` optimal (la complexité qui minimise l'erreur estimée) pour cet arbre de décision dont la lecture a été dégradée faute à un sur-apprentissage.

```
print(ptitanic.Arbre$cptable[which.min(ptitanic.Arbre$cptable[,4]),1])
```

Une fois le `cp` optimal identifié, il peuvent procéder à l'élagage de l'arbre avec le `cp` optimal obtenu.

```
ptitanic.Arbre_Opt <- prune(ptitanic.Arbre,cp=ptitanic.Arbre$cptable[which.min(
  ptitanic.Arbre$cptable[,4]),1])
```

ils ont réussi à obtenir un arbre “optimisé”.

4. Evaluation

1. Exécution du Rmd

L'exécution du RMarkdown se réalise parfaitement.

2. Qualité de la rédaction

Je trouve qu'il s'agit d'un bon travail. Les auteurs ont fourni un rapport bien organisé et facile à lire.

3. Qualité et lisibilité du RMarkdown

Le RMarkdown est globalement bien écrit, lisible et aéré.

4. Aspect didactique du dossier

La lecture de ce dossier est dans son ensemble aisée et accessible au plus grand nombre, même aux néophytes. Les descriptions.

Cependant, la partie “Test et Validation” est moins documentée.

5. Bibliographie

Il n'y a pas bibliographie.

5. Conclusion

Selon moi, il s'agit d'un bon travail.

En général, les deux auteurs ont bien documenté leur cas. C'est un travail bien construit, valorisé et didactique, avec un réel effort pour le rendre accessible au plus grand nombre.

Package lubridate

1. Entête

Nous allons évaluer le document ‘Tutoriel Lubridate’ élaboré par GASPARD Palay En cliquant ici vous trouverez le lien menant au GitHub: (<https://github.com/GaspardPalay/PSBX/tree/main/TutorielLubridate> (<https://github.com/GaspardPalay/PSBX/tree/main/TutorielLubridate>)).

2. Synthèse

Le package lubridate a été spécialement conçu pour gérer ces données de temps et les rendre facilement manipulables.

A noter que lubridate est disponible avec la collection de package TidyVerse en open source.

3. Extrait

L’auteur a réussi à mentionner des fonctions importantes de ce Package tel que : - La fonction `ceiling_date()` permet d’arrondir l’heure de la date à l’heure supérieure :

```
t <- ymd_hms("2020.11.09_17.56.32")
ceiling_date(t,"hour")
```

- `floor_date()` permet d’arrondir l’heure de la date à l’heure inférieure :

```
t <- ymd_hms("2020.11.09_17.56.32")
floor_date(t,"hour")
```

- `round_date()` permet d’arrondir l’heure de la date à l’heure la plus proche :

```
t <- ymd_hms("2020.11.09_17.56.32")
round_date(t,"hour")
```

L’auteur a également mentionner les fonctions qui permettent les calculs des périodes de temps écoulées ou de durée tel que:

-La fonction “ `diff` ” qui nous renseigne sur la “différence de temps” entre t1 et t2.

– `as.duration(diff)` affiche la différence de temps en secondes.

- `as.period(diff)` affiche la différence de temps en jours, en heures, en minutes et en secondes.

4. Evaluation

1. Exécution du Rmd

Le fichier Rmd ne s’exécute pas sans erreur sur la ligne 145.

2. Qualité de la rédaction

Je trouve qu’il s’agit d’un bon travail. Cependant l’auteur aurait dû afficher sur le document les exécutions de son code pour mieux expliquer son travail.

3. Qualité et lisibilité du RMarkdown

Le fichier est globalement bien écrit, lisible et aéré.

4.Aspect didactique du dossier

Le document est clair et simple .je trouve que les descriptions des fonctions sont compréhensibles et détaillées .

5. Bibliographie

Il n y'a pas de bibliographie.

5. Conclusion

Je trouve que c'est généralement un bon travail. L'auteur a fournit un document facile à lire et recherché. Cependant? il aurait dû ajouter les exécutions de son code pour mieux détailler et expliquer les fonctions.

#Package GGplot2

1. Entête

Nous allons évaluer le document 'Ggplot2' élaboré par Nicolas Allix & Marko Arsic.

En cliquant ici vous trouverez le lien menant au GitHub:

https://github.com/ARSICMrk/ARSIC_PSBx/tree/main/R_packages/Ggplot2

(https://github.com/ARSICMrk/ARSIC_PSBx/tree/main/R_packages/Ggplot2).

2. Synthèse

Le Ggplot2 est un package de visualisation utilisé aussi bien en R qu'en python. On l'utilise pour faire des graphiques à partir d'une base de données, mais on peut s'en servir pour bien plus que cela.

ggplot2 permet de visualiser des cartes, et de les rendre plus attractives.

Les auteurs ont choisi d'utiliser ce package pour créer une carte, pour cela ils ont aussi utilisé les packages map et data map .

3. Extrait

Les auteurs ont réussi à installer les différents packages tel que: ggplot2, maps, mapdata, dplyr et veridis, afin de manipuler les cartes .

Ensuite, ils ont expliqué la différence entre la carte du monde avec et sans le package de Ggplot2.

En effet, le package rend la carte plus esthétique et permet la sélection des pays qui nous intéressent.

4. Evaluation

1. Exécution du Rmd

Il n'existe pas un fichier Rmarkdown.

2. Qualité de la rédaction

Je trouve qu'il s'agit d'un assez bon travail. Les auteurs auraient dû mentionner et expliquer les différentes fonctions du package et afficher sur le document les exécutions du code pour mieux expliquer leur travail.

3. Qualité et lisibilité du RMarkdown

Le fichier PDF et md est globalement lisible. Cependant tout le document a la même police ce qui le rend mal organisé .

4. Aspect didactique du dossier

Le document est très simple . Il n'y a pas des explications des fonctions du package.

5. Bibliographie

Il n'y a pas de bibliographie.

5. Conclusion

Je trouve que c'est généralement un assez bon travail. L'auteur a fourni un document simple . Cependant il aurait dû ajouter les fonctions de son package et les exécutions de son code pour mieux détailler et expliquer son travail.

1. Entête

Le travail évalué dans ce document concerne le package JANITOR fait par Jiayue Liu et Olfa LAMTI.

En cliquant ici vous trouverez le lien menant au GitHub:

<https://github.com/OlfaLmt/PSBX/blob/main/Janitor/JANITOR%20DOCUMENTATION.pdf>

(<https://github.com/OlfaLmt/PSBX/blob/main/Janitor/JANITOR%20DOCUMENTATION.pdf>)

2. Synthèse

Le Package JANITOR a des fonctions simples pour examiner et nettoyer les données de façon simple et rapide . Il a été conçu pour les utilisateurs débutants et intermédiaires de R et est optimisé pour sa convivialité.

Les principales fonctions du concierge:

-formater parfaitement les noms de colonne data.frame;

-créer et formater des tableaux de fréquences à une, deux ou trois variables - pensez à un tableau amélioré ();

-fournir d'autres outils pour nettoyer et examiner les data.frames;

Les fonctions de tabulation et de rapport se rapprochent des fonctionnalités populaires de SPSS et Microsoft Excel.

3. Extrait

En effet Olfa et Jiayue n'ont pas mentionné le jeu de données avec laquelle ils ont travaillé. Cependant, en se basant sur leur rapport on va citer les différentes fonctions de ce Package :

-Les :: nous aide ici à appeler la fonction clean_names du package Janitor. Cette fonction génère un tableau avec les titres des colonnes dans le même format.

- tabyl() nous permet d'étudier les fréquences d'un vecteur d'une data frame.

- remove_empty() nous permet d'enlever les valeurs manquantes d'une data frame.

- get_dups() permet de récupérer les doublons d'une data Frame.

4. Evaluation

1. Exécution du Rmd

Il n'existe pas un fichier Rmarkdown.

2. Qualité de la rédaction

Je trouve qu'il s'agit d'un bon travail. Les auteurs ont fourni un rapport simple et facile à lire.

3. Qualité et lisibilité du RMarkdown

Le fichier PDF est globalement bien écrit, lisible et aéré .

4.Aspect didactique du dossier

Le document est très clair et simple .je trouve que les descriptions des fonctions sont compréhensibles bien détaillées .

5. Bibliographie

Il n'y a pas de bibliographie.

5. Conclusion

Je trouve que c'est globalement un bon travail. Les auteurs ont fourni un document clair et facile à comprendre.

1. Entête

Nous allons évaluer le document ***Package RMiner : Data Mining Classification et Méthode de Régression*** élaboré par Nina ZOUMANIGUI.

En cliquant ici vous trouverez le lien menant au GitHub: <https://github.com/Nina809/PSBX>
(<https://github.com/Nina809/PSBX>)

2. Synthèse

Le package Rminer aide à la manipulation des données algorithmiques pour le data mining concernant la classification et la régression (incluant la prévision de series temporelles).

3. Extrait

Le package est focalisé sur les fonctions suivantes du package Rminer qui peuvent être réparties dans 3 phases dans l'analyse des données :

- Data preparation : CasesSeries, delevels
- Modeling : fit, predict, mining
- Evaluation : mmetric, mgraph, mining

1. Data preparation : CasesSeries, delevels

- CasesSeries : créer un data frame d'une série temporelle en utilisant la fenêtre glissante.
- Delevels : réduire, remplace ou transforme les niveaux du data frame et la variable du facteur.

2. Modeling : fit, predict

- Fit : la fonction ajuste un modèle sélectionné de jeux de données et peut automatiquement ajuster les hyperparamètres.
- Predict : la fonction donne un modèle ajusté et calcule les prédictions pour un jeu de données.

3. Evaluation : mmetric, mgraph, mining

- Mmetric : fonction qui calcule les erreurs métriques de classification ou régression.
- Mgraph : fonction graphique
- Mining : la fonction réalise plusieurs exécutions ajustées et de prédictions, selon une méthode de validation et donne un nombre d'exécutions.

4. Evaluation

1. Exécution du Rmd

Il n'existe pas un fichier Rmarkdown.

2. Qualité de la rédaction

Je trouve qu'il s'agit d'un bon travail. L'auteur a fournit un rapport simple et facile à lire.

3. Qualité et lisibilité du RMarkdown

Le fichier est globalement bien écrit, lisible et aéré. .

4.Aspect didactique du dossier

Le document est clair et simple .je trouve que les descriptions des fonctions sont compréhensibles et détaillées .

5. Bibliographie

Il n y'a pas de bibliographie.

5. Conclusion

Je trouve que c'est généralement un très bon travail. L'auteur a fournit un document recherché et détaillé .

Package KScorrect“Auto-évaluation”

Ce travail permet au lecteur d’avoir une vision complète sur le Package Kscorrect et surtout sur le domaine de statistique,j’ai essaye de mentionner plusieurs fonctions importantes en expliquant leurs paramètres et utilité.L'exemple donné illustre de façon claire le concept de ce package.