

Les 12 Travaux d'AstéRix

Claire Mazzucato

29/01/2021

- Mes 5 critères permettant d'évaluer les travaux :
- Partie 1 : Les travaux basés sur R :
 - Travail N°1 : PALAY Gaspard et son Tutoriel Lubridate ;
 - Synthèse du travail
 - Commentaire des parties significatives
 - Evaluation selon mes critères
 - Conclusion
 - Travail N°2 : REN Claude et son Etude ggplot2 ;
 - Synthèse du travail
 - Commentaire des parties significatives
 - Evaluation selon mes critères
 - Conclusion
 - Travail N°3 : AUFRERE Thuy et son Etude RMiner ;
 - Synthèse du travail
 - Commentaire des parties significatives
 - Evaluation selon mes critères
 - Conclusion
 - Travail N°4 : Siva et Maxime et leur Etude Rpart ;
 - Synthèse du travail
 - Commentaire des parties significatives
 - Evaluation selon mes critères
 - Conclusion
 - Travail N°5 : REN Claude et son Etude Statsbomb ;
 - Synthèse du travail
 - Commentaire des parties significatives
 - Evaluation selon mes critères
 - Conclusion
- Partie 2 : Les travaux sur les mathématiques pour le Big Data:
 - Travail N°6 : LUTZ Rindra & ALIX Nicolas et leur travail sur les Arbres de Décision ;
 - Synthèse du travail
 - Commentaire des parties significatives
 - Evaluation selon mes critères
 - Conclusion
 - Travail N°7 : LIU Jiayue et son travail sur R-INLA ;
 - Synthèse du travail
 - Commentaire des parties significatives
 - Evaluation selon mes critères
 - Conclusion
 - Travail N°8 : ALLAKERE HORMO Maxime et son travail Régression sur une variable fonctionnelle ;
 - Synthèse du travail
 - Commentaire des parties significatives
 - Evaluation selon mes critères
 - Conclusion
 - Travail N°9 : BRETONNIERE Corentin, SERREAU Antoine, GUIGON Benjamin et leur travail sur les arbres de décisions ;
 - Synthèse du travail
 - Commentaire des parties significatives
 - Evaluation selon mes critères
 - Conclusion
 - Travail N°10 : William & Marko et leur travail sur la Cryptographie
 - Synthèse du travail
 - Commentaire des parties significatives
 - Evaluation selon mes critères
 - Conclusion
- Partie 3 : Auto-evaluation objective

Mes 5 critères permettant d'évaluer les travaux :

Ci-dessous, les 5 critères déterminés afin d'évaluer les travaux de mes camarades de promo:

- Critère 1: L'aspect visuel du travail (titre, police, paragraphe, etc)
- Critère 2: Utilisation de graphique, image, visuel, formules mathématiques avec Latex, bibliographie etc.
- Critère 3: Clarté concernant l'explication des travaux et bon fonctionnement des codes
- Critère 4: Qualité d'expression et d'orthographe
- Critère 5: Capacité à vulgariser des notions mathématiques ou R

Partie 1 : Les travaux basés sur R :

Travail N°1 : PALAY Gaspard et son Tutoriel Lubridate

(<https://github.com/GaspardPalay/PSBX/blob/main/TutorielLubridate/TutorielLubridate.pdf>) ;

Synthèse du travail

Dans son tutrotial, Gaspard nous montre à l'aide de différents exemples comment utiliser le package **lubridate**. Dans un premier temps il donne une explication générale du package. Celui-ci a pour objectif de convertir des éléments en format de date afin de pouvoir tirer profit de ce format pour la réalisation d'analyses, de création de graphiques et enfin de réaliser des calculs. Gaspard nous explique comment installer le package **lubridate** via la collection du package **TidyVerse**. La suite du tutoriel traite des aspects suivants: transformation d'une chaîne de caractères en date, manipulation des heures/dates en arrondissant des dates et en calculant des périodes de temps, les calculs arithmétiques, et enfin la notion d'heure POSIX ou Unix.

Commentaire des parties significatives

La partie que je considère la plus plus pertinente de son tutoriel concerne la partie sur le calcul arithmétique sur des périodes ou des durées. Au delà de l'intérêt de convertir des caractères en dates, **lubridate** permet de déterminer une nouvelle date en ajoutant des jours, semaines et mois à une date initiale à l'aide du code suivant:

```
install.packages("tidyverse")
```

ou simplement le package

```
install.packages("lubridate")
```

```
t1+months(9) # t1 + 9 mois
t1+ddays(287) # t1 + exactement 287 jours
ddays(287)/dweeks(1) # combien de semaines (exactement) pour 287 jours?
t2-dweeks(7) # t2 - 7 semaines
```

Ainsi, il est possible d'obtenir de nouvelles dates grâce aux fonctions expliquées par Gaspard. Dans ce tutoriel, il est aussi expliqué comment calculer des intervalles de temps comme ci-après.

```
## [1] 2016-12-27 05:45:00 GMT--2017-12-29 00:30:00 GMT
```

Evaluation selon mes critères

Globalement, ce tutoriel est très bien expliqué et l'illustration avec les codes R permet la bonne compréhension du package **lubridate**. Il remplit mon critère 1 car il y a la présence de titres, de paragraphes espacés et d'une police lisible. Gaspard n'a pas choisi d'illustrer son tutoriel avec des images ou visuels mais cela n'a pas d'impact sur la bonne compréhension du package. Enfin, l'expression et l'orthographe sont de bonne qualité et Gaspard a réussi à "vulgariser" des notions de R pour permettre à tous, y compris des débutants, de comprendre l'utilisation et la manipulation du package **lubridate**. Les seuls axes d'améliorations seraient, selon moi, d'afficher les résultats des codes R exprimés dans le tutoriel et enfin d'ajouter une bibliographie pour nommer les sources utilisées.

Critère 1 : L'aspect visuel du travail: 3/4

Critère 2 : Diversité du contenu: 2/4

Critère 3 : Clarté de l'explication et bon fonctionnement des codes: 3/4

Critère 4 : Qualité d'expression et d'orthographe: 4/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 4/4

Note Globale: 16/20.

Conclusion

Pour conclure, ce travail a été bien réalisé. Ne connaissant pas ce package, j'ai pu comprendre les bases de **lubridate** et son utilisation.

Travail N°2 : REN Claude et son Etude ggplot2

(https://github.com/Cldren/REN_PSBx/blob/main/Packages_presentation/Packages_presentation.pdf) ;

Synthèse du travail

Dans ce tutoriel, Claude nous montre l'utilisation du package **ggplot2**. Ce tutoriel est écrit en anglais et se compose d'une introduction, de parties sur l'utilisation d'alternatives et de **ggplot2**, une étude de cas, la personnalisation et se termine par une brève conclusion. Il commence par nous expliquer à quoi sert **ggplot2**. Il s'agit en effet d'une célèbre librairie de représentation graphique. Puis il nous explique qu'à titre d'exemple une base de données autour de modèles de voitures sera utilisée.

Tout d'abord, il nous explique l'utilisation d'alternative à **ggplot2** telle que **base plotting system** permettant la réalisation de représentations graphiques grâce à une syntaxe simple et minimaliste. Il nous explique ensuite l'utilisation de **Lattice** qui nous offre des résultats et une customization plus approfondies. Pour chaque solution alternative, Claude nous donne les avantages et inconvénients de leur utilisation. Enfin, **ggplot2** est expliqué dans un second temps. Ce qu'il faut retenir est que **ggplot2** est l'un des packages les plus utilisés sur R, cela est notamment dû au fait que ce package propose beaucoup de possibilités. Cela peut être complexe mais cela permet d'obtenir des graphiques de très bonne qualité.

Enfin Claude nous explique comment personnaliser les éléments de texte, de légende et de fond de graphique grâce à différentes fonctions.

Commentaire des parties significatives

La première partie de code intéressante dans ce document est celle-ci :

Ici, Claude nous donne, grâce à l'utilisation de la base de données liées aux modèles de voitures, une représentation graphique. En outre, il nous indique que quelque chose est anormal. Cela nous permet d'éveiller notre curiosité et de se poser différentes questions concernant ce graphique.

Ensuite, avec le code R suivant, il nous explique l'élément anormal. En effet, dans le premier graphique Claude n'a pas défini de séries temporelles, **ggplot2** permet la représentation graphique dans un cadre temporel.

Grâce aux explications concernant la syntaxe notamment `%%`, `seq()`, `as.Date()` rajouté après ce chunk, Claude nous permet de comprendre l'utilisation de **ggplot2** dans un exemple concret.

Evaluation selon mes critères

En ce qui concerne mon critère 1, Claude a soigné l'aspect visuel du tutoriel en y ajoutant des graphiques, des images et des couleurs utilisées. L'unique détail que je peux lui reprocher est le manque de grands titres pour structurer son tutoriel. En dehors de cela Claude a rendu le document agréable à lire. L'utilisation d'exemples concrets permet la bonne compréhension du package **ggplot2**. Ce que j'ai beaucoup apprécié dans son tutoriel est la mention "PROS" & "CONS" qui permet en quelques instants de comprendre les avantages et inconvénients de l'utilisation de packages liées aux représentations graphiques.

Critère 1 : L'aspect visuel du travail: 4/4

Critère 2 : Diversité du contenu: 3/4

Critère 3 : Clareté de l'explication et bon fonctionnement des codes: 3/4

Critère 4 : Qualité d'expression et d'orthographe: 4/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 2/4

Note Globale: 16/20.

Conclusion

Pour conclure, Claude a bien expliqué l'utilité de **ggplot2** notamment grâce à l'utilisation d'exemples concrets et l'explication complète de la syntaxe utilisée.

Travail N°3 : AUFRERE Thuy et son Etude RMiner (<https://github.com/T-AUF/PSBX/blob/main/Package%20RMiner.pdf>) ;

Synthèse du travail

Dans son tutoriel, Thuy nous présente le package **RMiner** qui a pour but de manipuler des données algorithmiques pour le Data Mining. Dans un premier temps Thuy nous donne des instructions pour pouvoir installer le package. Ensuite, elle définit dans une table une liste de fonctions pertinentes et nous invite à consulter la documentation dédiée. Ce tutoriel est divisé en 3 parties. Premièrement, Thuy nous explique la préparation des données grâce aux fonctions `casesSeries` et `delevels`. Puis, Thuy nous montre comment faire du Modeling grâce à l'utilisation des fonctions `predict` et `fit`. Enfin, le tutoriel présente comment faire une évaluation.

Commentaire des parties significatives

Une des parties que j'ai trouvée particulièrement pertinente est l'utilisation du code suivant permettant de réduire, remplacer ou transformer les niveaux du data frame et la variable du facteur

Evaluation selon mes critères

Thuy a réussi à "vulgariser" le package **RMiner** ce qui a permis une bonne compréhension du tutoriel car j'ignorais l'utilisation de ce package. En termes de visuel, le document de Thuy est assez minimaliste mais est facile à lire grâce aux différents titres ajoutés. Comme axe d'amélioration je noterais qu'il est pertinent d'ajouter une bibliographie afin de connaître les sources utilisées et d'illustrer son explication avec un exemple concret. Hormis ces détails, je considère que le travail de Thuy sur ce package a été bien réalisé.

Critère 1 : L'aspect visuel du travail: 2/4

Critère 2 : Diversité du contenu: 3/4

Critère 3 : Clareté de l'explication et bon fonctionnement des codes: 2/4

Critère 4 : Qualité d'expression et d'orthographe: 4/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 4/4

Note Globale: 15/20.

Conclusion

Pour conclure, ce document nous donne des informations utiles pour l'utilisation du package **RMiner** et peut être facilement compris de tous.

Travail N°4 : Siva et Maxime et leur Etude Rpart
(<https://github.com/Siva-chane/PSBX/blob/main/package%20Rpart/Rpart-package.pdf>)
;

Synthèse du travail

Dans ce tutoriel, mes deux camarades de promo nous donne une explication du package Rpart qui a pour rôle de construire des modèles de classification ou de régression d'une structure. Ils commencent par nous donner plus d'informations générales sur le fonctionnement de l'arbre de décision, les avantages et inconvénients de son utilisation. Ensuite, Siva et Maxime vont faire utiliser un exemple concret, à l'aide d'un dataset sur le naufrage du Titanic, pour nous montrer l'utilisation du package **Rpart**. En premier lieu, le tutoriel nous indique comment charger les librairies et charger les données. Puis, ils nous expliquent comment créer un dataset d'apprentissage et la construction de l'arbre de décision. Grâce à l'utilisation de représentations comme l'arbre de décision et également d'une courbe nous comprenons avec facilité comment fonctionne le package.

Commentaire des parties significatives

La partie que j'ai trouvé la plus pertinente concerne le chunk sur la construction de l'arbre de décision à partir de l'échantillon d'apprentissage car j'ignorais comment créer un arbre de décision sur R.

A l'aide de ce code, nous obtenons un arbre de décision assez complexe, et difficilement lisible pour réaliser une analyse. Siva et Maxime vont alors simplifier leur arbre grace au calcul du CP optimal (la complexité qui minimise l'erreur estimée)

Evaluation selon mes critères

L'aspect visuel de ce tutoriel est travaillé. En effet, il y a la présence de titres et de graphiques. L'ajout d'image et de couleurs pourraient apporter une touche visuelle encore plus sophistiquée mais cela n'a pas d'impact négatif sur la qualité du devoir. Siva et Maxime ont pu nous montrer à quoi sert un arbre de décision avec **Rpart** en utilisant des notions théoriques et en les appliquant à un exemple concret. Je pourrais reprocher à Siva et Maxime de ne pas avoir renseigné de bibliographie à leur tutoriel. La qualité d'expression et des explications permettent une bonne compréhension du package **Rpart**

Critère 1 : L'aspect visuel du travail: 3/4

Critère 2 : Diversité du contenu: 3/4

Critère 3 : Clarté de l'explication et bon fonctionnement des codes: 4/4

Critère 4 : Qualité d'expression et d'orthographe: 3/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 4/4

Note Globale: 17/20.

Conclusion

Globalement, ce travail a été très bien réalisé par Siva et Maxime. Grâce à leur tutoriel, j'ai pu apprendre comment installer le package, charger des données, créer un dataset d'apprentissage et enfin construire un arbre de décision.

Travail N°5 : REN Claude et son Etude Statsbomb
(https://github.com/Cldren/REN_PSBx/blob/main/Packages_presentation)
;

Synthèse du travail

Dans son tutoriel, Claude nous montre son usage de **Statsbomb**, un package dédié aux données gratuites de résultats au football. Il nous explique tout d'abord que l'on va être amenés à utiliser différents packages afin de réaliser les exercices du tutorial. Par la suite, il explique comment choisir un dataset. Il est à noter que Statsbomb offre un choix limité de jeux de données. Il nous donne pas à pas des indications avant de nous expliquer les fonctions utilisées. Après avoir choisi son jeu de données, il va créer une nouvelle table en prenant en compte la notion d'Expected Goals. Il va par la suite nous montrer qu'il est important de prendre en compte d'autres critères comme le nombre de passes pendant les jeux. Il va tout au long de ses explications illustrer ses exemples à l'aide de représentations graphiques réalisées grace à **ggplot2** jusqu'à arriver à une graphique assez complet sur deux équipes de football féminin Chelsea FCW et Arsenal WFC et se rapportant aux actions des joueuses sur une série temporelle précise (2018/2019).

Commentaire des parties significatives

Une des parties intéressante de son code est lorsqu'il génère le dernier graphique plus approfondi car il a utilisé une méthode en "entonnoir" pour nous montrer la représentation finale attendue.

Evaluation selon mes critères

En termes d'aspect visuel, le PDF de Claude est de bonne qualité. Tout d'abord, il a ajouté une page de garde, le logo de PSB est en en-tête de chaque page et il illustre son tutoriel à l'aide de différents graphiques ce qui rend sa lecture très agréable. Claude a écrit son tutoriel en anglais ce que je considère très inclusif et permet à un large public de comprendre ce tutoriel. Comme axe d'amélioration, je noterais le manque de phrases explicatives pour une meilleure interprétation des résultats obtenus à l'aide des graphiques. Outre ce détail, ce tutoriel est très bien structuré et a permis de manière globale une bonne compréhension du package **Statsboomb**

Critère 1 : L'aspect visuel du travail: 4/4

Critère 2 : Diversité du contenu: 3/4

Critère 3 : Clarté de l'explication et bon fonctionnement des codes: 4/4

Critère 4 : Qualité d'expression et d'orthographe: 4/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 2/4

Note Globale: 17/20.

Conclusion

Pour conclure, le travail de Claude est pertinent car son explication est facile de compréhension pour un public de tout niveau.

Partie 2 : Les travaux sur les mathématiques pour le Big Data:

Travail N°6 : LUTZ Rindra & ALIX Nicolas et leur travail sur les Arbres de Décision (<https://github.com/Nicolas-all/PSB1/blob/main/Arbres-de-D%C3%A9cision.pdf>) ;

Synthèse du travail

Le travail de Rindra et Nicolas porte sur la notion d'arbre de décision. En effet, elle est beaucoup utilisée dans l'exploration des données et en prise de décisions. Ce travail est composé de diverses parties composées d'une explication générale de l'arbre de décision, d'une brève partie liée à son histoire, à la construction d'un arbre de décisions et d'un arbre de classification. Ensuite, ils mettent en évidence les avantages et limites à l'utilisation d'arbres de décisions. A l'aide d'un exemple simple, Rindra et Nicolas nous explique comment construire un arbre de décision.

Commentaire des parties significatives

La partie que j'ai trouvée la plus intéressante concerne celle sur les arbres de classifications. En effet, j'avais déjà quelques connaissances sur l'arbre de décision, précédemment vu lors de l'étude du travail de Siva et Maxime sur le package **Rpart** mais j'ignorais que l'on pouvait traiter des données qualitatives à l'aide d'un arbre de classification.

Ils expliquent que dans le cas d'un arbre de classifications, chaque noeud est associé une coupure, et à chaque coupure, une variable de coupure X_{jt} selon laquelle on découpera le noeud.

Pour qu'un noeud soit pertinent, le seuil de coupure doit avoir une valeur qui maximise un caractère. On note ce seuil de coupure α_t :

- Si $X_{jt}^{it} \in A_t$ alors on continue dans la branche de gauche
- Si $X_{jt}^{it} \notin A_t$ alors on continue dans la branche de droite

contrairement à l'arbre de décision:

- Si $X_{jt}^{it} \leq \alpha_t$ alors on continue dans la branche de gauche
- Si $X_{jt}^{it} > \alpha_t$ alors on continue dans la branche de droite

Cette explication nous permet de comprendre comment fonctionne un arbre de classification pour l'exploitation de données qualitatives et quelles sont les différences avec un arbre de décision.

Evaluation selon mes critères

D'un point de vue aspect visuel du devoir, mon critère 1 et 2 ont été respectés car ils ont utilisé des titres pour structurer leur travail. Par ailleurs le format du devoir est diversifié. En effet, Rindra et Nicolas ont utilisé des représentations graphiques pour illustrer leur travail. Pour ce qui concerne les explications et la clarté des notions abordées tout le long de ce RMD est de qualité. Ce que j'ai notamment apprécié est leur capacité à résumer des aspects mathématiques complexes pour que ce soit compréhensible de tous. Même si cela n'a pas de conséquence sur la bonne compréhension du travail, une bibliographie aurait pu être ajoutée à ce devoir.

Critère 1 : L'aspect visuel du travail: 3/4

Critère 2 : Diversité du contenu: 3/4

Critère 3 : Clarté de l'explication et bon fonctionnement des codes: 4/4

Critère 4 : Qualité d'expression et d'orthographe: 3/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 3/4

Note Globale: 16/20.

Conclusion

Pour conclure, ce travail permet d'avoir des bonnes bases pour comprendre ce que représente un arbre de décision et de classification dans le cadre de l'analyse de données qualitatives et quantitatives.

Travail N°7 : LIU Jiayue et son travail sur R-INLA (https://github.com/liu-jiayue/psbx/blob/main/travaux_math/Models_spatio_temporels.pdf) ;

Synthèse du travail

Dans son devoir, Jiayue nous apporte beaucoup d'informations sur les bases de "R-INLA" ainsi que son utilité. Par ailleurs il nous donne une approche mathématique avec la probabilité, la modélisation et inférence bayésienne.

Jiayue nous explique le concept de l'inférence bayésienne. Celle-ci détermine qu'à partir d'événements connus on peut calculer la probabilité des différentes causes hypothétiques. Contrairement à la probabilité classique, la pensée bayésienne se caractérise par la tolérance à la subjectivité.

Jiayue nous explique ensuite l'utilisation du package **R-INLA** pour construire un modèle spatio-temporel. Il nous explique comment installer le package car celui-ci ne fait pas parti de CRAN et nécessite une installation manuelle.

Enfin, avec l'utilisation d'un jeu de données sur le nombre de personnes victime de cancer des poumons, Jiayue nous permet d'analyser le risque relatif du cancer en fonction de critères géographiques et temporelles.

Commentaire des parties significatives

La partie que j'ai trouvé la plus pertinente est la démonstration du modèle Bernardinelli afin d'estimer le risque relatif du cancer de poumon avec l'équation suivante :

$$Y_{ij} \sim Po(E_{ij}\theta_{ij})$$

$$\log(\theta_{ij}) = \alpha + u_i + v_i + (\beta + \delta_i) * t_j.$$

où pour la commune i en l'an j , Y_{ij} est le nombre de cas de cancer observé E_{ij} est le nombre de cas de cancer théorique, et θ_{ij} s'agit du risque relatif (RR). L'intersection dénotée par α , ($u_i + v_i$) qui capte l'effet aléatoire d'une commune, β qui représente l'effet de tendance linéaire. Par ailleurs δ_i représente le décalage entre cette tendance générale et la tendance spécifique à une commune. Enfin, t_j capte l'aspect temporel.

Evaluation selon mes critères

Pour ce qui concerne l'aspect visuel, le travail de Jiayue est très bien structuré, il y a des titres, des images, une table des matières et le logo de PSB a même été ajouté. Parmi tous les travaux étudiés, ce travail est l'un des mieux présenté. Cela rend la lecture de son travail vraiment agréable. L'expression et l'orthographe sont de qualité et l'explication de Jiayue est adaptée au niveau de tous.

Critère 1 : L'aspect visuel du travail: 4/4

Critère 2 : Diversité du contenu: 4/4

Critère 3 : Clarté de l'explication et bon fonctionnement des codes: 4/4

Critère 4 : Qualité d'expression et d'orthographe: 4/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 3/4

Note Globale: 19/20.

Conclusion

Pour conclure, ce document est très bien structuré et l'explication de Jiayue est très complète. J'ai pu, grâce à l'utilisation d'exemple et à l'explication de formules mathématiques comprendre l'intérêt de manipuler des données avec **R-INLA**

Travail N°8 : ALLAKERE HORMO Maxime et son travail Régression sur une variable fonctionnelle (https://github.com/mallaker/PSB_X/blob/main/Dossier%20Maths/REGR) ;

Synthèse du travail

Dans son devoir, Maxime nous présente un sujet autour de la régression sur une variable fonctionnelle. Il aborde d'abord des problématiques liées à la régression dans différents domaines tels que la technologie avec la reconnaissance vocale, les sciences avec le secteur de la génétique ou encore le domaine de l'économétrie. Maxime va ensuite nous présenter les méthodes d'analyse de données fonctionnelles. Il nous donne les paramètres essentiels à la réalisation d'une régression comme l'importance de représenter graphiquement les données avec une courbe et le fait que les résidus doivent suivre une loi statistique.

Commentaire des parties significatives

La dernière partie de ce document est celle que j'ai trouvée la plus pertinente. En effet, Maxime nous explique avec clarté les modèles de régressions sur variable fonctionnelle au travers des estimateurs paramétriques et non-paramétriques. Il nous présente l'équation suivante:

1. $Y = aX + b + \epsilon, X \sim N(\mu, \gamma^2)$ et $N(0, \sigma^2)$
2. $Y = aX + b + \epsilon, E(\epsilon|X) = 0$
3. $Y = r(X) + \epsilon, r \in C(\mathbb{R})$ et $E(\epsilon|X) = 0$

Maxime nous explique la différence entre chaque équation avec l'équation (1) qui est une équation paramétrique et l'équation (3) non-paramétrique. L'équation (2) est quant à elle une équation paramétrique et non paramétrique selon les deux points de vue suivant : - Non-Paramétrique si la loi du couple (X, Y) est supposée appartenir à un espace indexé par un nombre fini de paramètres réels. - Paramétrique si l'opérateur de régression r est supposé appartenir à un espace indexé par un nombre fini de paramètres réels.

Evaluation selon mes critères

Les critères qui concernent la clarté de l'explication est totalement respecté. En effet, Maxime a su exprimer l'utilité de la notion abordée dans ce document. En matière d'aspect visuel, le document est assez minimaliste mais très bien structuré ce qui permet une bonne lisibilité. Les formules mathématiques sont bien présentées, ce qui permet de les comprendre dans le détail. Si je devais lui recommander des axes d'amélioration ce serait d'ajouter plus de diversité concernant le format du document, comme ajouter des images, graphiques, etc.

Critère 1 : L'aspect visuel du travail: 3/4

Critère 2 : Diversité du contenu: 2/4

Critère 3 : Clarté de l'explication et bon fonctionnement des codes: 4/4

Critère 4 : Qualité d'expression et d'orthographe: 4/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 3/4

Note Globale: 16/20.

Conclusion

Pour conclure, le travail de Maxime a été très bien réalisé même si son format pourrait être légèrement amélioré. Outre ce détail, le travail de Maxime est de qualité.

Travail N°9 : BRETONNIERE Corentin, SERREAU Antoine, GUIGON Benjamin et leur travail sur les arbres de décisions
https://github.com/T-Hak/PSBX/blob/main/Tutoriel%20pracma/gr02_hakam_tarik_pracma.pdf
 ;

Synthèse du travail

Antoine, Corentin et Benjamin présentent dans ce travail les arbres de décisions utilisés pour la prédiction. Ils expliquent d'abord en détail les arbres de régressions avec la notion de pureté. En effet, un nœud peut être pur ou impure en fonction de si tous les individus associés à une valeur appartiennent à cette classe. Grâce à l'utilisation d'une représentation graphique nous pouvons remarquer facilement les différences entre un nœud pur et impur. Il expliquent ensuite que la pureté d'un nœud est calculée grâce à l'indice de Gini. Plus l'indice se rapproche de 0 plus le nœud est pur. Par la suite, mes camarades de promo expliquent la notion de coût d'un nœud. Celui-ci calcule à quel point le choix de la variable de décision est bonne. Puis, à l'aide d'un exemple réalisé avec **Iris**, ils nous présentent un arbre de décision visuel avec le résultat de la prédiction. Enfin, ils expliquent assez brièvement la notion d'arbre de classification permettant l'exploitation de données qualitatives.

Commentaire des parties significatives

L'une des parties que j'ai trouvées les plus pertinentes sont l'explication de l'indice de Gini pour déterminer si un nœud est pur ou non.

$$G_i = 1 - \sum_{k=1}^n P_{ik}^2$$

D'autre part, j'ai apprécié l'exemple utilisé avec **Iris** sur R car cela permet de comprendre en détail comment il est possible de construire un arbre de régression et quel est le résultat de prédiction obtenu.

Evaluation selon mes critères

Ce travail respecte les critères liés à l'aspect visuel. En effet la présence de titres, images, représentations permet la bonne lecture du document. La notion d'arbre de décision est expliqué avec clarté et le travail est bien structuré. Les graphiques et expressions LaTeX sont de qualité Au niveau de la qualité. La seule remarque que je peux faire et l'oubli de bibliographie.

Critère 1 : L'aspect visuel du travail: 3/4

Critère 2 : Diversité du contenu: 3/4

Critère 3 : Clareté de l'explication et bon fonctionnement des codes: 4/4

Critère 4 : Qualité d'expression et d'orthographe: 3/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 3/4

Note Globale: 16/20.

Conclusion

Pour conclure, Antoine, Corentin et Benjamin on réalisé un travail de qualité et a permis d'en connaitre davantage sur l'utilisation d'arbres de décisions pour l'exploitation de données qualitatives et quantitatives.

Travail N°10 : William & Marko et leur travail sur la Cryptographie (https://github.com/ARSICMrk/ARSIC_PSBx/blob/main/Maths_BD/Cryptographie)

Synthèse du travail

Dans ce travail, William et Marko nous initie à la notion de Cryptographie. Elle a d'abord vu le jour dans le domaine militaire mais s'est développé à de nombreux secteurs d'activité par la suite. La Cryptographie est notamment utilisée pour garantir la confidentialité des échanges sur Internet et pour protéger des données dites sensibles. Dans ce travail, mes camarade de promo nous explique plusieurs types de cryptages: le cryptage à clé symétrique avec le codage César et Vigenère, et le cryptage à clé asymétrique avec des fonctions à sens unique, le principe de chiffrement, de signature, etc. Ce travail se termine par une ouverture sur l'informatique quantique.

Commentaire des parties significatives

La partie que j'ai trouvé particulièrement intéressante est leur explication du principe de chiffrement car ils définissent de manière claire chaque élément qui compose la formule mathématique suivante.

$$f_D(M) = f_D[f_C(m)] = m$$

Ainsi, toute personne interceptant un message crypté ne peut le lire. Seul la personne utilisant la fonction de décryptage peut permettre de lire le message.

Evaluation selon mes critères

J'ai trouvé ce travail passionnant car la notion de Cryptographie éveille une certaine curiosité et William et Marko ont pu expliquer de manière claire et complète cette dernière. L'aspect visuel du travail pourrait être légèrement amélioré, en effet, les titres se différencient peu entre eux et le sommaire a été oublié. L'aspect minimaliste du devoir ne pose pas de problème à sa compréhension.

Critère 1 : L'aspect visuel du travail: 2/4

Critère 2 : Diversité du contenu: 2/4

Critère 3 : Clareté de l'explication et bon fonctionnement des codes: 4/4

Critère 4 : Qualité d'expression et d'orthographe: 4/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 4/4

Note Globale: 16/20.

Conclusion

Globalement, ce travail sur la Cryptographie a été très bien réalisé. En effet, même en ayant peu de connaissances sur le sujet, il est facile de comprendre. Marko et William pourraient améliorer l'aspect du devoir pour un meilleur confort de lecture. Mise à part ce détail, le travail est de qualité.

Partie 3 : Auto-evaluation objective

L'étude de ces 10 derniers travaux m'a permis de comparer mes devoirs par rapport à ceux des autres et à remarquer des lacunes ou qualités dans le travail réalisé.

A travers cette auto-evaluation, je jugerai mon devoir avec les critères définies plus haut concernant l'aspect visuel de mon devoir, la diversité du contenu, la clareté de l'explication et le bon fonctionnement des codes, la qualité d'expression et la capacité à vulgariser des notions complexes.

Mon tutoriel Network3D (<https://github.com/clairemazzucato/PSBX/blob/main/Packages/NetworkD3/NetworkD3.pdf>)

Mon tutoriel concernant **Network3D** permet de connaître le fonctionnement de ce package et ses différentes utilisations. J'explique avec plus ou moins de précision les différents types de graphiques que l'on peut utiliser comme le `simple network`, le `force network` le `sanky network` et enfin le `radial network`. En revanche, je n'apporte pas assez d'exemples concrets ce qui rend le devoir assez descriptif.

L'aspect visuel de mon document est correct mais nécessiterait un sommaire, un meilleur alignement des titres. En revanche, comparativement avec d'autres devoir j'ai utilisé plusieurs format dans ce tutoriel comme des images, et des polices différentes. D'autres part, j'ai intégré une bibliographie qui mériterait un format plus approfondi.

Je considère que mon devoir est assez simple de compréhension et sa lecture est facile. En revanche je pourrais le rendre plus sophistiqué.

Critère 1 : L'aspect visuel du travail: 3/4

Critère 2 : Diversité du contenu: 3/4

Critère 3 : Clareté de l'explication et bon fonctionnement des codes: 3/4

Critère 4 : Qualité d'expression et d'orthographe: 3/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 3/4

Note Globale: 15/20.

Mon travail sur la Prédiction en IA

(https://github.com/clairemazzucato/PSBX/blob/main/Mathe%CC%81matiques%20pour%20le%20Big%20Data/Dossier_Maths.pdf)

Ce travail s'appuie sur l'étude de 3 papiers de thèses autour de la prédiction en Intelligence Artificielle. En faisant une comparaison avec les autres travaux de mes camarades, je me suis rendue compte que nous étions le seul groupe à avoir réellement respecté les consignes. En effet, nous avons analysé 3 papiers de thèses comme demandé alors que les autres membres de la promo faisait une analyse d'un sujet de mathématiques.

En termes de visuel, je considère que nous avons fait un travail de qualité. En effet, notre travail se compose d'un sommaire, de titres qui structurent très bien nos propos, d'un logo de PSB, de la présence d'annexes, etc.

L'explication des formules mathématiques sont assez évasives. En effet, s'agissant de sujets assez complexes et au vu de mon niveau de mathématiques je ne me sentais pas à l'aise à l'explication de celles-ci.

L'explication globales des notions abordées et la qualité d'expression rend la lecture de ce travail assez facile, à mon sens.

Critère 1 : L'aspect visuel du travail: 4/4

Critère 2 : Diversité du contenu: 2/4

Critère 3 : Clareté de l'explication et bon fonctionnement des codes: 3/4

Critère 4 : Qualité d'expression et d'orthographe: 3/4

Critère 5 : Capacité à vulgariser des notions mathématiques ou R: 2/4

Note Globale: 14/20.

Pour conclure, mes 2 travaux ont été réalisé dans le respect des consignes. J'aurais pu apporter plus de détails et plus d'exemples pour illustrer mes propos mais j'estime qu'ils restent compréhensible pour un public de tous niveaux.