

Retour d'expérience sur l'enseignement de la science des données biologiques en classe inversée

Guyliann Engels 1*

Philippe Grosjean 2†

Résumé

R est enseigné dans les cours de science des données du cursus en biologie de l'Université de Mons, en Belgique. Cinq cours en classe inversée et apprentissage par projet sont dispensés sur les quatre dernières années de la formation (du deuxième bachelier à la seconde années de Master). Nous utilisons le package `{learnitdown}` afin de gérer nos différents outils d'apprentissage `{bookdown}`, `h5p`, `{learnr}`, `{shiny}` et les projets individuels ou de groupes sur GitHub. L'ensemble de ces activités sont analysées afin d'obtenir un suivi précis en temps réel de l'évolution de chaque étudiant. Ces derniers peuvent également consulter un rapport de progression individuelle à tout moment. Cet exposé s'intéresse l'apprentissage des étudiants analysé au travers de ces données.

Mots-clefs : Science des données biologiques – suivi de l'apprentissage – classe inversée – `learnitdown`

Développement

R est enseigné dans le cursus de biologie de l'Université de Mons en Belgique. Depuis 2018, les cours classiques de biostatistique ont été remplacés par des cours de sciences des données biologiques en classe inversée. Ces nouveaux cours ont permis d'intégrer des nouvelles notions liées à la reproductibilité, au données ouvertes ou encore à l'Open Science. Ce changement d'approche pédagogique a été accompagné de l'utilisation progressive de nouveaux outils comme des `{bookdown}` intégrés dans un site Wordpress, des applications `{shiny}`, des tutoriels `{learnR}`,...

Les activités d'apprentissages sont découpées en trois niveaux. D'abord, les étudiants lisent les notions théoriques illustrées dans un contexte biologique grâce à un cours en ligne écrit avec `{bookdown}` et répondent à des questions intégrées (H5P ou `{Shiny}`). Ensuite, les étudiants réalisent des tutoriels `{learnR}` afin de se familiariser avec le code R associé à ces notions théoriques. Pour finir, ils réalisent des travaux individuels ou de groupe afin d'appliquer les notions apprises sur des cas pratiques en biologie (projets). Les travaux sont hébergés sur GitHub dans une organisation (<https://github.com/BioDataScience-Course>) et générés via GitHub Classroom.

Sur trois ans, le nombre d'étudiants suivant le cours de science des données biologique I a augmenté de 60%. De ce fait, nous utilisons le package R `{learnitdown}` pour gérer et enregistrer efficacement la progression individuelle de l'ensemble des apprenants au sein d'une plateforme d'apprentissage intégrée (*LMS Learning Management Systems*) centrée autour de R et RStudio.

Les résultats des activités H5P, `{shiny}` et `{learnr}` sont directement collectés dans une base de données. Les projets sont analysés grâce au gestionnaire de version git. Sur l'année académique 2020-2021, les 133 étudiants répartis sur trois années, du Bachelier 2 au Master 1, ont réalisé plus de 8850 commits et ont modifié au total plus de 310.000 lignes dans leurs projets. On dénombre également plus de 300.000 entrées dans la base de données associée aux activités H5P, `{learnr}` et `{shiny}`. A noter que les étudiants visualisent à tout moment leur rapport de progression et que leur évaluation est uniquement basée sur ces exercices. Ces cours en classe inversée et par projets motivent les étudiants en biologie qui ne sont pas forcément tous enclins au départ à apprécier la science des données et travail sur ordinateur. Plus de 90% des activités sont réalisées systématiquement par les étudiants si l'on excepte le profil des étudiants en décrochage. D'ailleurs,

*Service d'écologie numérique, Institut Complexys & Infortech, Université de Mons, Belgique, guyliann.engels@umons.ac.be

†Service d'écologie numérique, Institut Complexys & Infortech, Université de Mons, Belgique, philippe.grosjean@umons.ac.be

le taux d'abandon est très faible avec moins de 10% pour le premier cours en second Bachelier, qui est aussi le plus sélectif.

L'analyse transversale des données sur les trois années montre que la note finale obtenue est positivement corrélée à l'assiduité. Plusieurs profils d'étudiants se dégagent, allant des étudiants qui abandonnent aux étudiants perfectionnistes. Le pourcentage de réussite est stable et toujours supérieur à 80%. Les activités d'apprentissages les plus difficiles sont détectées via le nombre de tentatives et le temps nécessaires pour chaque exercice. Les erreurs récurrentes sont aussi repérables dans la base de données qui enregistre chaque réponse à chaque exercice. Le matériel pédagogique est adapté chaque année afin de corriger les défauts ainsi détectés. Par exemple, le nombre de projets a augmenté d'année en année et ils sont répartis en assignations individuelles fortement cadrées d'une part, et des assignations plus complexes et ouvertes à réaliser en groupe, d'autre part. Globalement, le nombre de commits par étudiant sur l'ensemble des projets reste stable, mais cet étagement permet un apprentissage plus progressif qui s'avère bénéfique. Les deux premières années, nous observons que les notes obtenues pour les différents projets sont très fortement corrélées avec les notes obtenues aux examens en fin d'année. De ce fait, nous avons supprimé les examens de fin d'années pour laisser plus de temps aux étudiants de travailler sur leurs projets. L'évaluation ne se fait plus qu'au travers des exercices et de ces projets.