

Étude des techniques de prédiction en Machine Learning pour lutter contre l'échec et le décrochage scolaire.

Énoncé :

Au cours de ces dernières années, un certain nombre de projets de recherche et d'initiatives ont vu le jour pour s'attaquer à la question du suivi des performances académiques des élèves (Pistilli et Arnold, 2010 ; Romero et al., 2008 ; Smith et al, 2012). En général, les chercheurs utilisent des algorithmes d'apprentissage (par exemple, les **modèles linéaires**, les **réseaux bayésiens**, les **machines à vecteurs de support** ou **séparateurs à vaste marge « SVM »** et les **arbres de décision**), pour développer des modèles prédictifs qui tentent d'identifier les élèves à risque académique (risque d'échec ou de décrochage).

Par exemple, on s'intéresse à détecter, relativement tôt dans le semestre, les étudiants qui éprouvent des difficultés académiques durant un cours en exploitant leurs données (Jayaprakash et al., 2014). Afin de d'augmenter la précision des prédictions et de réduire leur variabilité, on introduit les méthodes d'ensemble ou « **stacking** » qui consistent à combiner des modèles prédictifs en les appliquant sur des versions modifiées des données (**bagging** ou **boosting**) (Lauría & al, 2018). On s'intéresse aussi aux méthodes de détection d'anomalies (Lauría, E. J. (2021), dans le cas où l'effectif d'élèves à risque, est très réduit comparé à ceux qui réussissent.

Travail à réaliser :

Réaliser une étude bibliographique sur les travaux existants s'intéressant à la prédiction de l'échec scolaire, afin de ressortir les méthodes de prédictions les plus utilisées dans ce domaine, de connaître leur potentiel et leurs limites et lister les attributs qui sont principalement utiles pour ce genre de prédiction.

Cette étude servira de préparation pour l'élaboration d'une solution de prédiction sur des données scolaires.

N.B :

Le décrochage scolaire c'est environ 98 000¹ jeunes qui sortent annuellement du système de formation en France. Comme conséquences du décrochage on évalue à 50% le taux de chômage d'un jeune sorti du système scolaire, et à 230 000 euros les coûts associés à ce décrochage, cumulés tout au long de sa vie. Pour les jeunes concernés, cette situation est source de difficultés sociales et économiques majeures.

Références :

1. Pistilli, M.D., Arnold, K.E., 2010. Purdue Signals: Mining Real-time Academic Data to Enhance Student Success. Campus 22–24.
2. Romero, C., Ventura, S., Garcia, E., 2008. Data mining in course management systems: Moodle case study and tutorial. Comput Educ 51, 368–384.
3. Smith, V.C., Lange, A., Huston, D.R., 2012. Predictive Modeling to Forecast Student Outcomes and Drive Effective Interventions in Online Community College Courses. J. Asynchronous Learn. Netw. 16, 51–61.
4. Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J. R., Baron, J. D., 2014. Early Alert of Academically At-Risk Students: An Open Source
5. Analytics Initiative. J. Learn. Anal. 1, 42.
6. Lauría, E. J., Presutti, E., Kapogiannis, M., & Kamath, A. (2018, March). Stacking Classifiers for Early Detection of Students at Risk. In *CSEdu (1)* (pp. 390-397).

¹ <https://www.gouvernement.fr/ce-qu-il-faut-savoir-sur-le-decrochage-scolaire>

7. Lauría, E. J. (2021). Framing Early Alert of Struggling Students as an Anomaly Detection Problem: An Exploration. In *CSEDU (1)* (pp. 26-35)

Contacts :

Fahima DJELIL. Maître de Conférence, Lab-STICC UMR CNRS 6285/[Équipe MOTEL](#). IMT Atlantique Bretagne Pays de Loire, Department Informatique. Technopôle Brest-Iroise CS 29 238 Brest.

Email: Fahima.djelil@imt-atlantique.fr

tél: +332 29 00 14 32 /+336 25 15 17 06

Serge GARLATTI. Professeur Émérite. Lab-STICC UMR CNRS 6285/[Équipe MOTEL](#). IMT Atlantique Bretagne Pays de Loire, Department Informatique. Technopôle Brest-Iroise CS 29 238 Brest.

tél: +33 2 29 00 14 53

Email: serge.garlatti@imt-atlantique.fr