



HAL
open science

Vers une Approche Éthique pour la Prédiction Automatique du Risque d'Échec Scolaire ★

Fahima Djelil, Jean-Marie Gilliot

► **To cite this version:**

Fahima Djelil, Jean-Marie Gilliot. Vers une Approche Éthique pour la Prédiction Automatique du Risque d'Échec Scolaire ★. Environnements Informatiques pour l'Apprentissage Humain et Intelligence Artificielle (EIAH & IA), Jun 2022, Saint-Étienne, France. hal-03784588

HAL Id: hal-03784588

<https://hal.science/hal-03784588>

Submitted on 23 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vers une Approche Éthique pour la Prédiction Automatique du Risque d'Échec Scolaire^{*}

Fahima Djelil¹[0000-0001-8449-2062] et Jean-Marie Gilliot¹[0000-0003-4569-3552]

IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France
fahima.djelil@imt-atlantique.fr
jm.gilliot@imt-atlantique.fr

Résumé Cet article donne un bref état de l'art sur la prédiction d'apprenants à risque d'échec scolaire, et pose la question d'éthique dans la prédiction. Une rapide revue de littérature nous a permis de ressortir les attributs de prédiction les plus fréquemment utilisés. Ces attributs étant le plus souvent démographiques et donc sensibles, ont un réel potentiel sur la précision de prédiction algorithmique mais peuvent engendrer des résultats discriminants vis à vis de certaines populations sous-représentées. L'équité algorithmique, un sujet actuellement en plein essor en machine learning, montre des pistes de recherche prometteuses pour atténuer les biais algorithmiques et offrir des solutions éthiques d'aide à la décision.

Keywords: Échec scolaire · Analytique Prédicative · Équité.

1 Introduction

Le décrochage scolaire est un problème social majeur dans toutes les sociétés du monde. En France, c'est environ 98 000 jeunes qui sortent annuellement du système de formation [1]. 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. Les recherches suggèrent que la décision d'abandonner l'école est le résultat d'une longue évolution et d'un cumul de frustrations engendrées par les échecs scolaires et par les difficultés relationnelles avec les pairs, les enseignants et les parents [7]. Les efforts de prévention doivent être déployés le plus tôt possible pendant que le jeune est encore à l'école car il est plus facile de le maintenir à l'école que de l'amener à y retourner. Il est donc important pour les professionnels de l'éducation de mieux cerner les raisons du décrochage afin de l'éviter [7].

*. Soutien du Labex CominLabs (ANR-10-LABX-07-01) et de la Direction du Numérique pour l'Éducation (DNE), dans le cadre de ses Groupes Thématiques Numériques (GTNum) 2020-2022 - GTNum8 IA_EO : impact de l'Intelligence Artificielle à travers l'Éducation Ouverte.

Au cours de ces dernières années, un certain nombre de recherches et d’initiatives ont vu le jour pour s’attaquer à la question du suivi automatique des performances académiques des élèves [21,22,24]. On s’appuie alors sur des algorithmes d’apprentissage automatique (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 [10].

Une des préoccupations majeure des travaux existants sur la prédiction d’échec scolaire consiste à maximiser la précision et la performance algorithmique. Par exemple, afin 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) [18]. On s’intéresse aussi aux méthodes de détection d’anomalies [17], dans le cas où l’effectif d’élèves à risque est très réduit comparé à ceux qui réussissent. Il existe également plusieurs revues systématiques, mais se limitent souvent à lister les techniques de prédiction utilisées et les attributs de prédiction les plus fréquents (e.g. [15,2]). Les questions d’éthique dans le recueil et le traitement des données sont rarement abordées.

Cet article donne un bref état de l’art sur la prédiction automatique d’échec scolaire, et s’intéresse particulièrement aux questions éthiques dans la prédiction. Dans ce qui suit, nous listons dans un premier temps les attributs de prédiction les plus fréquents (facteurs de décrochage). Ensuite, nous abordons les questions éthiques dans le traitement de données pour la prédiction d’apprenants à risque d’échec ou d’abandon scolaire. Enfin, nous concluons par des orientations de recherche.

2 Attributs de Prédiction

Les attributs utilisés pour la prédiction de la performance des apprenants sont académiques, sociaux, démographiques, personnels et familiaux (Tableau 1). Selon le contexte et le pays où les données sont recueillies, on peut choisir des marqueurs tels que l’âge, le sexe, le lieu de résidence mais aussi des marqueurs sociaux tels que la profession des parents, la situation familiale ou le revenu [16]. Cependant, dans la plupart des cas, ce sont les résultats scolaires qui ressortent comme les principaux facteurs qui affectent les résultats des apprenants et donc leur échec (notes d’examens, évaluations, informations sur le cursus) [15].

3 Éthique et Équité

L’éthique en analytique prédictive cherche à définir ce qui est vulnérable et dommageable, afin de prévenir les préjudices potentiels envers les apprenants, mais aussi envers toutes les parties prenantes [23]. Cette définition implique :

Tableau 1. Revue de différents attributs utilisés pour la prédiction du risque d'échec scolaire, selon différents contextes.

Référence	Attributs	Contexte
1. [5]	Résultats académiques et informations de cursus	Cycle 1 universitaire en Science, Technologie, Ingénierie et Mathématiques (Brésil, 2010)
2. [11]	Informations personnelles (nom, date de naissance, adresse, nationalité, sexe et tuteur légal), besoins éducatifs spécifiques et difficultés d'apprentissage (retards, handicaps, rapports psychoéducatifs et médicaux), mauvais comportement (incident, sanction, enseignant ayant appliqué la sanction, date de l'incident, avertissement individuel ou collectif), nombre d'inscriptions ou redoublement.	Niveau secondaire - Lycée (Espagne, 2005-2007)
3. [19]	Résultats académiques et informations démographiques (lieu de résidence, nationalité, sexe)	Niveau universitaire (Macédoine du Nord, 2011-2014)
4. [3]	Résultats académiques	Cycle 1 universitaire, étude de technologie (Arabie Saoudite, 2013-2014)
5. [16]	Informations démographiques et sociales (sexe, pauvreté, environnement scolaire, inadéquation des méthodes d'enseignement)	Niveau secondaire (Inde, 2013)
6. [25]	Résultats académiques, information de cursus, situation professionnelle, intérêt pour la formation, activités extra-scolaires, et informations démographiques (âge, sexe, lieu de résidence, situation maritale, niveau d'éducation des parents)	Cycle 1 universitaire, étude de santé (Iran, 2013)
7. [17]	Informations démographiques (âge, sexe) et d'aptitude, informations sur le cursus, notes obtenues aux devoirs, projets et examens et toute autre activité contribuant à la note finale de l'étudiant au cours des six premières semaines du semestre enregistrées dans le LMS.	Cycle 1 universitaire (États-Unis, 2018-2020)

- la question de discrimination implicite ou explicite, par laquelle un apprenant reçoit ou ne reçoit pas un soutien, sur la base de ce qui pourrait être vu de l’extérieur comme une caractéristique personnelle quelconque,
- l’étiquetage, où les apprenants peuvent être étiquetés en fonction d’une combinaison de caractéristiques et traits, potentiellement pour la durée de leurs études, comme étant différents des autres,
- et la validité du traitement spécifique de groupes d’apprenants sur la base d’hypothèses concernant leurs caractéristiques communes.

L’inclusion de données protégées dans les modèles de prédiction, peuvent engendrer des discriminations de groupes d’étudiants minoritaires et accentuer les inégalités déjà existantes [6]. Ce sont des données à caractère personnel comprenant des traits ou des caractéristiques sur la base desquels la discrimination est illégale. On parle alors d’attributs protégés [27], ou d’informations sensibles [9]. Par ailleurs, l’utilisation explicite de données socio-démographiques peut donner lieu à des prédictions plus précises et améliorer l’efficacité des interventions basées sur des décisions algorithmiques [20].

Les biais et les stéréotypes sont considérés comme ayant un rôle important dans l’analyse prédictive [23]. Pourtant, le biais renvoie à la tendance des individus à avantager les membres d’une catégorie perçue comme possédant un haut statut par rapport aux membres de l’autre catégorie perçue comme dotée d’un bas statut [4]. Plus précisément, ce biais peut être relatif aux stéréotypes et aux préjugés, mais peut aussi représenter des discriminations qui sont des comportements défavorables envers un individu en raison de son appartenance à une catégorie [4]. On recommande de mettre en avant les biais et les stéréotypes dans les tentatives institutionnelles visant à catégoriser les profils d’apprenants à risque, afin de montrer que les analyses prédictives initiales concernant le potentiel de réussite d’un groupe d’individus sont fausses ou incomplètes [23]. En déterminant ce qui pourrait constituer une vulnérabilité dans le contexte des learning analytics, les institutions doivent s’assurer que les analyses sont effectuées sur des jeux de données robustes et suffisamment représentatifs [23].

Plus récemment, *l’équité algorithmique* a émergé comme champ de recherche pour lequel il existe un intérêt croissant, en particulier dans le domaine d’apprentissage automatique en éducation [14]. La notion *d’équité* est fondée sur l’existence de groupes et visent à garantir que certains groupes ne sont pas défavorisés et reçoivent un traitement égal [12]. En ce sens, [8] a introduit le concept *d’égalité des chances*, qui exige que le résultat prédit par un classifieur doit être indépendant d’un attribut protégé. Cela implique de collecter des caractéristiques qui capturent plus directement l’objectif cible, indépendamment de sa corrélation avec l’attribut protégé.

Plusieurs recherches émergentes s’intéressent à l’équité algorithmique pour la prédiction d’échec scolaire (e.g. [27,13,26]). L’étude présentée dans [27], montre la construction de modèles d’apprentissage automatique pour prédire l’abandon des études après une année universitaire et compare la performance globale et l’équité des prédictions du modèle avec ou sans quatre attributs protégés (sexe, minorité sous représentée, étudiant de première génération, i.e. dont les

parents ne possèdent aucun diplôme d'enseignement supérieur, et besoins financiers élevés). Par exemple, un modèle de prédiction qui atteint la même précision sur les élèves garçons et filles est considéré comme équitable en termes de précision. [13] propose de prédire les étudiants à risque tout en évitant les biais discriminatoires potentiels. Cela comprend l'évaluation de l'équité algorithmique des modèles dans différents groupes en fonction de plusieurs métriques et l'application d'une méthode d'atténuation des biais afin de définir la parité pour les sous-groupes présentant des résultats inéquitables. [26] propose un framework pour détecter et documenter les risques d'iniquité lors de la prédiction d'apprenants à risque de décrochage.

4 Conclusion

Les attributs de prédiction d'apprenants à risque peuvent présenter des caractéristiques très sensibles de l'individu (informations socio-économiques, familiales, personnelles ...). De plus certains attributs peuvent comporter des informations biaisées et stéréotypées. L'inclusion de ces attributs dans les algorithmes de prédiction est donc sujette à des questions éthiques. Alors que certains travaux sur la prédiction d'échec scolaire omettent les question éthiques et se focalisent sur la performance algorithmique, certains d'autres travaux se préoccupent au contraire de l'éthique dans le traitement de données personnelles et éducatives des apprenants, et soulignent les préjudices que cela peut causer.

L'éthique en prédiction d'échec scolaire est un sujet en pleine émergence et une manière de traiter les différentes questions qui en découlent consiste à étudier l'équité algorithmique des modèles de prédiction. L'équité algorithmique en analytique prédictive est un sujet à part entière et son application au problème d'éthique dans la prédiction d'échec scolaire est très récent. Tout l'enjeu est alors de savoir si un jeu de données contient des risques d'iniquité, d'assurer la performance algorithmique tout en assurant l'équité et d'atténuer des résultats de prédiction pouvant engendrer des discriminations. L'ensemble de ces questions présentent de réelles opportunités de recherche.

Références

1. Ce qu'il faut savoir sur le « décrochage scolaire » (2016), <https://www.gouvernement.fr/ce-qu-il-faut-savoir-sur-le-decrochage-scolaire>
2. Albreiki, B., Zaki, N., Alashwal, H. : A systematic literature review of student performance prediction using machine learning techniques. *Education Sciences* **11**(9), 552 (2021)
3. Altujjar, Y., Altamimi, W., Al-Turaiki, I., Al-Razgan, M. : Predicting critical courses affecting students performance : a case study. *Procedia Computer Science* **82**, 65–71 (2016)
4. André, A., Louvet, B. : Influence de la structuration de l'apprentissage coopératif sur les relations filles-garçons en éducation physique. *Revue des sciences de l'éducation* **40**(1), 83–105 (2014)

5. Barbosa Manhães, L.M., da Cruz, S.M.S., Zimbrão, G. : Towards automatic prediction of student performance in stem undergraduate degree programs. In : Proceedings of the 30th Annual ACM Symposium on Applied Computing. pp. 247–253 (2015)
6. Barocas, S., Hardt, M., Narayanan, A. : Fairness and machine learning. fairmlbook.org. URL : <http://www.fairmlbook.org> (2019)
7. Fortin, L., Royer, É., Potvin, P., Marcotte, D., Yergeau, É. : La prediction du risque de décrochage scolaire au secondaire : facteurs personnels, familiaux et scolaires. Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement **36**(3), 219 (2004)
8. Hardt, M., Price, E., Srebro, N. : Equality of opportunity in supervised learning. Advances in neural information processing systems **29** (2016)
9. Hu, Q., Rangwala, H. : Towards fair educational data mining : A case study on detecting at-risk students. International Educational Data Mining Society (2020)
10. Jayaprakash, S.M., Moody, E.W., Lauría, E.J., Regan, J.R., Baron, J.D. : Early alert of academically at-risk students : An open source analytics initiative. Journal of Learning Analytics **1**(1), 6–47 (2014)
11. Jiménez-Gómez, M.Á., Luna, J.M., Romero, C., Ventura, S. : Discovering clues to avoid middle school failure at early stages. In : Proceedings of the Fifth International Conference on Learning Analytics and Knowledge. pp. 300–304 (2015)
12. Joseph, M., Kearns, M., Morgenstern, J., Neel, S., Roth, A. : Fair algorithms for infinite and contextual bandits. arXiv preprint arXiv :1610.09559 (2016)
13. Karimi-Haghighi, M., Castillo, C., Hernandez-Leo, D., Oliver, V.M. : Predicting early dropout : Calibration and algorithmic fairness considerations. arXiv preprint arXiv :2103.09068 (2021)
14. Kizilcec, R.F., Lee, H. : Algorithmic fairness in education. arXiv preprint arXiv :2007.05443 (2020)
15. Kumar, M., Singh, A., Handa, D. : Literature survey on educational dropout prediction. International Journal of Education and Management Engineering **7**(2), 8 (2017)
16. Kumar, P.S., Panda, A.K., Jena, D. : Mining the factors affecting the high school dropouts in rural areas. International Journal of Advanced Computer Engineering and Communication Technology (IJACECT), ISSN (Print) pp. 2278–5140 (2013)
17. Lauría, E.J. : Framing early alert of struggling students as an anomaly detection problem : An exploration. In : CSEDU (1). pp. 26–35 (2021)
18. Lauría, E.J., Presutti, E., Kapogiannis, M., Kamath, A. : Stacking classifiers for early detection of students at risk. In : CSEDU (1). pp. 390–397 (2018)
19. Nikolovski, V., Mishkovski, I., Stojanov, R., Chorbev, I. : Educational data mining : Case study for predicting student dropout in higher education. In : Proceedings of the 12th International Conference on Informatics and Information Technologies (2015)
20. Paquette, L., Ocumpaugh, J., Li, Z., Andres, A., Baker, R. : Who’s learning? using demographics in edm research. Journal of Educational Data Mining **12**(3), 1–30 (2020)
21. Pistilli, M.D., Arnold, K.E. : Purdue signals : Mining real-time academic data to enhance student success. About Campus **15**(3), 22–24 (2010)

22. Romero, C., Ventura, S., García, E. : Data mining in course management systems : Moodle case study and tutorial. *Computers & Education* **51**(1), 368–384 (2008)
23. Slade, S., Prinsloo, P. : Learning analytics : Ethical issues and dilemmas. *American Behavioral Scientist* **57**(10), 1510–1529 (2013)
24. Smith, V.C., Lange, A., Huston, D.R. : Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of asynchronous learning networks* **16**(3), 51–61 (2012)
25. Teshnizi, S.H., Ayatollahi, S.M.T. : A comparison of logistic regression model and artificial neural networks in predicting of student's academic failure. *Acta Informatica Medica* **23**(5), 296 (2015)
26. Vasquez Verdugo, J., Gitiaux, X., Ortega, C., Rangwala, H. : Faired : A systematic fairness analysis approach applied in a higher educational context. In : *LAK22 : 12th International Learning Analytics and Knowledge Conference*. pp. 271–281 (2022)
27. Yu, R., Lee, H., Kizilcec, R.F. : Should college dropout prediction models include protected attributes? In : *Proceedings of the Eighth ACM Conference on Learning@ Scale*. pp. 91–100 (2021)