



## Journée d'étude du GTnum LS2N #IA\_EO (17/06/21) : IA et prédiction du risque de décrochage scolaire

### Résumé

#### Résumé en français

Retour sur la journée d'étude du GTnum LS2N #IA\_EO (17/06/21), sous le format atelier, qui s'est tenue sur le thème de la prédiction du décrochage scolaire.

#### Résumé en anglais

Review of the GTnum LS2N #IA\_EO study day (17/06/21), in the workshop format, which was held on the theme of predicting school dropout.

#### Auteurs

Fahima Djelil, Maîtresse de Conférence à IMT Atlantique Bretagne Pays de Loire, Lab-Sticc, Brest / Nicolas Hernandez, Maître de conférences, Université de Nantes, LS2N / Serge Garlatti, Professeur émérite, IMT Atlantique Bretagne Pays de Loire, Lab-Sticc / Mélanie Pauly Harquevaux, Cheffe de projet international à l'Université de Nantes



GTNUM 8

#IA\_EO

INTELLIGENCE ARTIFICIELLE  
ET ÉDUCATION OUVERTE

**GTnum #IA\_EO Workshop**  
Digital Thematic Working Groups  
on Artificial Intelligence and  
Open Education

IMT Atlantique/Lab-STICC  
in partnership with  
University of Nantes/LS2N, MEN and DANE

**Machine Learning for Early Detection of Academically At-Risk Students:  
Challenges, Technologies and Opportunities for Research**

Eitel J. M. Lauría  
Professor of Information Technology & Systems  
School of Computer Science & Mathematics  
Marist College, Poughkeepsie, NY, U.S.A.

Marist College, Poughkeepsie, NY, U.S.A. **MARIST**



## Retour sur le programme

- **14h-15h** : Conférence donnée par le professeur Eitel J. M. Lauría (Professor of Information Technology & Systems, School of Computer Science & Mathematics, Marist College, Poughkeepsie, NY, U.S.A), *"Machine Learning for Early Detection of Academically At-Risk Students : Challenges, Technologies and Opportunities for Research"*  
Website: <https://foxweb.marist.edu/users/jf4n/index.html>
- **15h-16h** : Ateliers et discussions en petits groupes
- **16h-16h30** : Restitution

## La conférence

La présentation a été réalisée par le professeur Eitel J.M. Lauría sous la thématique : *"Machine Learning for Early Detection of Academically At-Risk Students : Challenges, Technologies and Opportunities for Research"*.

Vous pouvez accéder à cette conférence en rediffusion [ici](#).

**GTnum #IA\_EO Workshop**  
Digital Thematic Working Groups  
on Artificial Intelligence and  
Open Education

IMT Atlantique/Lab-STICC  
in partnership with  
University of Nantes/LS2N, MEN and DANE

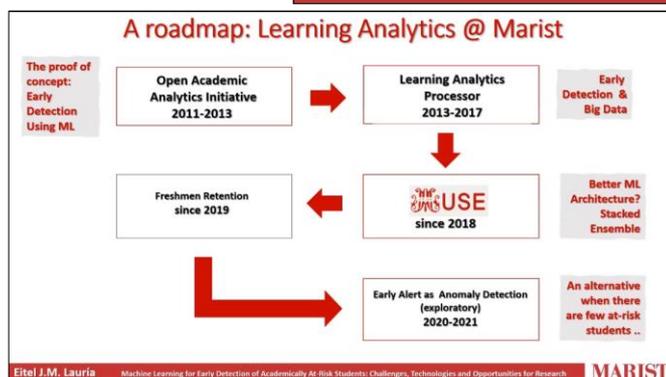
---

**Machine Learning for Early Detection of Academically At-Risk Students:  
Challenges, Technologies and Opportunities for Research**

Eitel J. M. Lauría  
Professor of Information Technology & Systems  
School of Computer Science & Mathematics  
Marist College, Poughkeepsie, NY, U.S.A



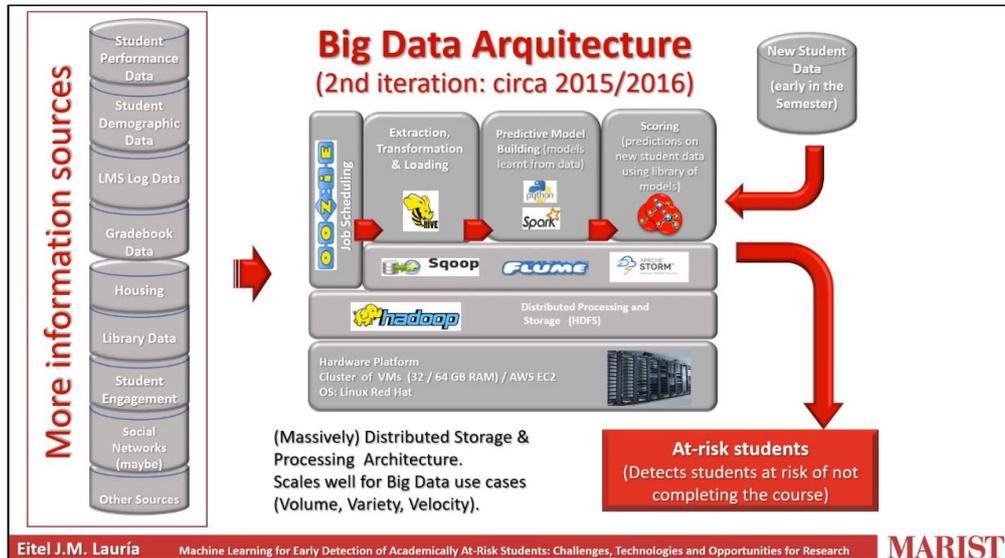

Marist College, Poughkeepsie, NY, U.S.A.



**Research Questions**

- **How good are predictive models ?**
- **What are good model predictors ?**
  - ✓ Typical predictors:
    - Student performance Data: standardized tests (SATs), current GPA
    - Student Demographic Data: age, gender, etc.
- **Can we add effort as a predictor ? How do we measure effort ?**
  - ✓ LMS Activity Log Data
- **Can we add partial grades as predictors ?**
  - ✓ Good grades remain as the best predictors of academic success (a tautology ?)
- **Are intervention strategies effective? Which ones ?**

Eitel J.M. Lauría Machine Learning for Early Detection of Academically At-Risk Students: Challenges, Technologies and Opportunities for Research MARIST



## Final Thoughts

### Open Issues and Future Research

- Data-driven vs. theory-driven analysis (or explain vs. predict: social science vs data science ? )
- Does learning analytics allow us to identify at-risk students that the typical instructor would miss ? (my answer is YES)
- Predictive power of LMS event log data
- In a heavily unbalanced setting, is anomaly detection a valid option ?
  - An anomaly is a rare event, but is a rare event an anomaly?
- Is transfer learning a valid tool in the toolbox for institutions that lack resources to collect the data ? How portable are predictive models ? Can we fine tune them?
- Ethical issues:
  - Other features ? Privacy vs big brother
  - Providing predictions about a student's potential performance has the potential to create an instructor expectation that may not benefit the student (self-fulfilling prophecy)
  - Informing the student about a risk of failure introduces issues related to withdrawal that have not been addressed by many.

Source : [https://mediaserver.univ-nantes.fr/videos/gtnum-ia\\_eo-journee-detudes-n2-workshop-du-17-juin-2021/](https://mediaserver.univ-nantes.fr/videos/gtnum-ia_eo-journee-detudes-n2-workshop-du-17-juin-2021/)



## Les ateliers et leur restitution

### Atelier N° 1 : traitement éthique des données scolaires

Animateur : Nicolas Hernandez, Maître de conférences, Université de Nantes, LS2N

Participants :

- Marine Roche, ATER, Université de Nantes, CREN
- Bastien Masse, Université de Nantes, LS2N
- Ilker Birkan, Université de Nantes
- Sébastien Iksal, Maître de conférences HDR, Université du Mans, LIUM
- Catherine Lemonnier, DANE, Académie de Nantes

Ce groupe s'est focalisé sur la question de l'éthique et ses enjeux pour l'accès et l'exploitation des données ainsi que pour l'open data et la recherche. Le groupe a d'abord pointé le manque d'ouverture des données des plateformes d'enseignement. Cette problématique, résultant à la fois des modèles économiques et juridiques des plateformes et du manque d'interfaces de programmation ouvertes vers les données, entraîne des enjeux d'accès, de collecte, de partage et d'exploitation de données pour la recherche et l'innovation en général.

A ces enjeux viennent ensuite se greffer la nécessité de réaliser un travail de sensibilisation à l'exploitation de ces données dans les milieux scolaires. Ce travail doit passer par une éducation commune et généralisée avec un double objectif : d'une part, assurer la transparence sur la manière dont on peut exploiter les données éducatives, tout en protégeant les individus et en respectant les règles de protection des données à caractère personnel (RGPD, souveraineté des données numériques,...) ; et d'autre part, promouvoir les bénéfices des applications qui pourraient être produites si l'accès aux données scolaires était davantage ouvert et libre. Cela permettra de limiter les craintes pouvant être ressenties dans les milieux scolaires au regard de l'exploitation des données.

Le dernier enjeu relevé est de nature universelle. Comme pour tout résultat de recherche ou pour toute innovation, il est important que chaque acteur de la chaîne conduise systématiquement des réflexions sur les effets possibles de l'exploitation des données scolaires (effets Golem et Pygmalion).

Du point de vue de l'IA, une des manières de procéder pour participer au travail de transparence est de se fonder sur des approches permettant l'interprétation et l'explication des analyses des données pour l'utilisateur final. En effet, l'explicabilité de l'IA constitue un véritable enjeu de recherche. Il est primordial de faciliter le transfert de la recherche auprès des acteurs du terrain.



## Atelier N° 2 : Prédiction du risque de décrochage scolaire

Animatrice : Fahima Djelil, Maître de Conférences, IMT Atlantique Bretagne Pays de Loire, Lab-Sticc

Participants :

- Eitel Lauria, Professeur, Marist College, Poughkeepsie, NY (USA)
- Colin de la Higuera, Professeur, titulaire de la Chaire UNESCO REL, Université de Nantes, LS2N
- Lucile Fouré-Bechade, Université de Nantes, LS2N
- Esther Félix, étudiante en stage M2, Université de Nantes, LS2N

Dans cet atelier, nous avons échangé en particulier avec notre invité, Pr. Eitel Lauria (USA). Nous avons abordé la question de la prédiction de l'échec scolaire sous un regard extérieur. En effet, la prédiction d'élèves à risque de décrochage scolaire a fait l'objet de nombreuses recherches empiriques aux Etats-Unis.

Au sein du Marist College, Poughkeepsie, NY (USA), il existe l'initiative *Open Academic Analytics* (circa 2011-2013), pour la prédiction de la réussite des étudiants par l'identification de modèles (*patterns*) dans des grands volumes de données, en employant des techniques d'apprentissage automatique. L'objectif étant de prédire les étudiants à risque académique dès les premières semaines de cours, puis d'intervenir afin d'améliorer la réussite.

D'autres initiatives ont suivi, comme « MUSE - Marist Universal Student Experience », depuis 2018. Ce projet a reçu plusieurs distinctions (Computerworld Honors Laureate and Finalist in the Emerging Technology category - *over 700 nominations* ; Campus Technology Innovator Award - *one of only 9 recipients over 230 applied*).

Le dispositif MUSE fournit un tableau de bord permettant la visualisation des prédictions de réussite académique des étudiants en premier cycle, dès les 6 premières semaines d'un cours s'étalant sur 15 semaines. Le tableau de bord offre des vues adaptées aux enseignants et aux étudiants. Il permet aux enseignants d'intervenir de manière aisée, par exemple en contactant directement les étudiants concernés ou en déclenchant des actions spécifiques auprès des services compétents de l'établissement.

Au sein de Marist College, le bénéfice de cette initiative se traduit par un pourcentage de 5% d'étudiants qui évitent le décrochage, ce qui représente un bon chiffre à l'échelle de cette école.

## Atelier N° 3 : La remédiation par l'IA

Animateur : Serge Garlatti, Professeur émérite, IMT Atlantique Bretagne Pays de Loire, Lab-Sticc

Participants :

- Jean-Marie Gilliot, IMT Atlantique, Lab-Sticc
- Issam Rebai, IMT Atlantique, Lab-Sticc
- Mélanie Pauly Harquevaux, Université de Nantes, LS2N
- Paul Hellwig, Inspecteur d'Académie, Académie de Lille
- Franck Poirier, Université de Bretagne Sud, Lab-Sticc
- Annie Foret, Université Rennes1, IRISA

Suite à la présentation du Pr. Eitel Lauria, ce groupe a réfléchi au processus de remédiation à partir d'analyse de données en utilisant les méthodes d'IA.



## Carnet Hypothèses « Education, numérique et recherche »

Direction du numérique pour l'éducation

TN2

La remédiation nécessite de faire du diagnostic, par exemple de l'évaluation formative pour intervenir au fur et à mesure de la progression des élèves dans les exercices. La remédiation peut intervenir à différents niveaux (globale dans le cursus, par discipline, par cours, par exercice, etc.). Une des manières possibles serait de faire des évaluations formatives pour identifier le « modèle cognitif » de l'apprenant. L'IA pourrait détecter les profils cognitifs pour ensuite s'en servir pour orienter les phases de remédiation.

Pour effectuer l'analyse des données des apprenants, la diversification des sources et de la nature des données à collecter permettrait d'augmenter la qualité de la prédiction et donc de la remédiation. En effet, cela pourrait mieux affiner les profils des apprenants, et donc l'identification des profils cognitifs.

Enfin, il est nécessaire de définir la place de l'humain dans la remédiation, et donc de distinguer ce qui relève des tâches de la machine et celle de l'humain. Il s'agit donc de définir le rôle de l'IA et celui de l'humain dans le processus de remédiation.



## Communication

#GTnum @Edu\_num #IA\_EO

[https://twitter.com/Chaire\\_RELIA](https://twitter.com/Chaire_RELIA)

<https://chaireunescorel.ls2n.fr/>

## Référence

Pauly Harquevaux, M. (2021, juillet 12). GTnum #IA\_EO / journée d'études n°2—Workshop du 17 juin 2021 : Conférence donnée par le professeur Eitel J. M. Lauría « Machine Learning for Early Detection of Academically At-Risk Students : Challenges, Technologies and Opportunities for Research ». Webtv de l'Université de Nantes.

<https://mediaserver.univ-nantes.fr/permalink/v1261ac51ad67a2phnov/>

## A consulter aussi

DNE-TN2. (2020, avril 21). Analyse des traces d'apprentissage – Learning analytics : Productions du GTnum 2 [Billet]. Éducation, numérique et recherche. <https://edunumrech.hypotheses.org/1559>

DNE-TN2. (2021a, mai 6). Point d'étape du #GTnum LS2N #IA\_EO (05/05/21) : Prédire la performance académique, découverte des travaux de Luis Galárraga [Billet]. Éducation, numérique et recherche. <https://edunumrech.hypotheses.org/2982>

DNE-TN2. (2021b, mai 10). Webinaire commun #GTnum #Scol\_IA #IA\_EO (26/05/21) : Outils et ressources pour se former à l'IA [Billet]. Éducation, numérique et recherche. <https://edunumrech.hypotheses.org/3004>

DNE-TN2. (2021c, juin 14). Webinaire du GTnum LS2N #IA\_EO (25/03/21) : Le projet ApLLy, test d'Auto-positionnement en Ligne des Lycées [Billet]. Éducation, numérique et recherche. <https://edunumrech.hypotheses.org/3101>

DNE-TN2. (2021d, juillet 6). Webinaire du GTnum LS2N #IA\_EO (15/04/21) : Etude sur les explications du résultat d'un classifieur [Billet]. Éducation, numérique et recherche. <https://edunumrech.hypotheses.org/3233>

Eitel J.M. Lauría, Information Technology & Systems. (s. d.). Marist College. Consulté 16 juillet 2021, à l'adresse <https://www.marist.edu/computer-science-math/faculty/eitel-lauria>

Galarraga, L. (2021, février 5). GTnum #IA\_EO / axe 1—Exploitation Intelligible des Learning Analytics—Webinaire 1. Webtv de l'Université de Nantes. <https://mediaserver.univ-nantes.fr/permalink/v12619a3299f5xrqaxiv/>

Guillelé, M. (2019). Extraction de connaissances interprétables dans des séries temporelles [Thèse de doctorat, Rennes 1]. <http://www.theses.fr/2019REN1S102>

Pauly Harquevaux, M. (2021, mai 26). GTnum #IA\_EO / axe 1—Exploitation Intelligible des Learning Analytics—Webinaire 2 Le projet ApLLy, test d'Auto-positionnement en Ligne des Lycées. Webtv de l'Université de Nantes. <https://mediaserver.univ-nantes.fr/permalink/v1261a13af8b6hsrfjihv/>