

ÉNAP

ÉCOLE NATIONALE
D'ADMINISTRATION
PUBLIQUE

enap.ca

Étude de cas sur la valorisation des données au Centre de services scolaire du Val-des-Cerfs

Par Christian Boudreau (Ph. D.)

Professeur à l'ENAP

Mai 2023

Table des matières

Introduction	2
1. La conception des premiers modèles de prédiction : 2017-2020	3
2. Le projet pilote : 2020 à 2021	7
3. L'après projet pilote : 2022 à 2023	12
4. Des conditions facilitantes	16
5. Des enjeux de valorisation	20
Conclusion	24
Bibliographie	26
Annexes	27

Introduction

Les modèles de prédiction sont une composante importante de l'intelligence artificielle. Ils permettent de prédire des comportements futurs à partir de l'analyse d'événements du passé. Ces modèles peuvent être utilisés par les organisations pour saisir des occasions d'affaires, notamment à partir de l'analyse de leur environnement et de la détection de pratiques ou de tendances émergentes. Les modèles de prédiction peuvent aussi servir à anticiper et à gérer des problèmes de toutes sortes, comme les crises financières, les crimes contre la personne, les risques de récidive, les cybermenaces, les incendies, les épidémies, les échecs scolaires et les catastrophes naturelles. L'éducation est un secteur propice au déploiement de modèles prédictifs, compte tenu des problèmes potentiels de décrochage, de redoublement ou d'échec scolaire auxquels doivent faire face les intervenants de ce secteur.

La présente étude de cas porte sur un projet de valorisation de données, initié en 2017 par le Centre de services scolaire du Val-des-Cerfs (ci-après CSSVDC) et son directeur général, visant notamment la conception et l'utilisation de modèles de prédiction du risque de décrochage. Tout en relatant l'histoire du projet, l'étude montre les avantages, les conditions facilitantes, les enjeux et les leçons qui accompagnent le déploiement de ces modèles en éducation. Les résultats de l'étude s'appuient principalement sur des entretiens semi-dirigés menés auprès de deux représentants du CSSVDC¹ et de deux experts externes² engagés dans le projet depuis le début. Ces entretiens ont été réalisés aux mois de décembre 2022 et janvier 2023. Les propos des répondants sont en italiques dans le texte. Ils sont tantôt insérés entre des guillemets (« »), tantôt mis en retrait du texte. Je tiens à remercier la disponibilité et la grande franchise des personnes rencontrées en entretien. Leurs propos constituent la matière première de l'étude.

¹ En particulier le directeur général, Eric Racine, et le directeur du Développement de solutions d'analytique avancée, Jérémie Alluard.

² En particulier Éric Marcotte et Hugo Juhel. Au moment de l'entretien, ces deux experts étaient les dirigeants d'une petite entreprise appelée « Sciance Inc. » Au début du projet, ils étaient des employés de la firme Raymond Chabot Grant Thornton.

1. La conception des premiers modèles de prédiction : 2017-2020

Depuis 2017, le centre de services scolaire du Val-des-Cerfs (ci-après CSSVDC) développe et expérimente, avec le soutien de deux experts externes, des modèles de prédiction pour détecter le plus tôt possible les élèves du secondaire qui présentent des signes annonciateurs d'un éventuel décrochage scolaire. Le directeur général du CSSVDC, Eric Racine, est à l'origine (et le principal instigateur) de ce projet d'expérimentation d'outils d'intelligence artificielle. « *Ayant été lui-même un décrocheur durant son parcours scolaire, Eric Racine connaît très bien le chemin qu'empruntent celles et ceux qui s'éloignent des bancs d'école.* » (CSSVDC, 2022a)

La persévérance scolaire a été une préoccupation constante tout au long de la carrière de cet ancien enseignant, devenu tour à tour directeur adjoint d'école, directeur d'école, puis directeur de centre de services scolaire. Cet intérêt s'est notamment traduit par le désir de mieux comprendre les causes du décrochage et d'en repérer les premiers symptômes, en particulier auprès d'élèves qui présentent peu ou pas de signes avant-coureurs. Alors que certains décrocheurs potentiels, aux prises avec des difficultés scolaires apparentes, peuvent être facilement détectés à l'avance par les professeurs et autres intervenants, puis être pris en charge par le milieu scolaire, d'autres décrocheurs potentiels, au parcours moins problématique, peuvent passer sous le radar des enseignants et des directions d'écoles, comme l'explique Eric Racine :

Lorsque je suis arrivé à la direction d'établissement, je ne comprenais pas que j'avais des élèves de secondaire 5 qui fonctionnaient pourtant assez bien... Ils n'avaient aucun échec, ils n'avaient pas beaucoup d'absences. Mais, du jour au lendemain, ils décrochaient et on ne comprenait pas pourquoi... Et peu importe les moyens que j'essayais de mettre en place pour les ramener sur les bancs d'école, dans bien des cas le décrochage s'était cristallisé, puis ils ne revenaient plus... Je les appelais nos élèves beiges... C'est ceux-là qui se confondent entre la peinture et le mur... Ils ne sont pas sur le radar, ils n'ont jamais été en échec... C'est ces élèves-là que j'essayais d'aller chercher.

Pour détecter le décrochage scolaire et intervenir avant qu'il ne survienne, Eric Racine a commencé à analyser les nombreuses données sur les élèves auxquelles il avait accès comme

directeur adjoint, puis comme directeur d'école, avec le soutien des enseignants. Cependant, ces analyses étaient ardues, compte tenu des difficultés de croiser les données pour retracer le parcours scolaire des élèves depuis la première année du primaire jusqu'au secondaire. On devait extraire les données dans différents fichiers et les croiser de façon manuelle.

Je rencontrais mes enseignants... et on regardait les résultats des élèves... On retournait en arrière dans les fichiers des élèves. On essayait de faire des graphiques sur leur progression en mathématiques, puis en français. Puis, on regardait au niveau des absences. Mais c'était tout du silo, puis c'était toute faite à la mitaine, un par un. (Eric Racine)

Au moment où Eric Racine obtenait le poste de directeur général à la Commission scolaire du Val-des-Cerfs³, en 2015, le taux de décrochage sur ce territoire était l'un des plus élevés au Québec. Au terme d'un cycle de restructuration budgétaire et d'optimisation des services avec Raymond Chabot Grant Thornton, mené en 2017, un membre de cette firme comptable⁴ offrait à Eric Racine (ci-après directeur général) de mettre à sa disposition deux jeunes experts⁵ en valorisation des données pour collaborer à un projet d'intelligence artificielle visant à contrer le décrochage scolaire sur le territoire du CSSVDC. Conscient du potentiel de valorisation des nombreuses données collectées dans son centre de services scolaire, le directeur général a accepté l'offre de la firme.

Plutôt que de faire des autopsies à la fin de l'année scolaire en vue de la prochaine année, il y aurait lieu de croiser les données dont on dispose pour être en mesure de prédire les élèves qui ont un plus grand potentiel de décrochage et de concentrer nos énergies et nos mesures sur ces élèves (Directeur général).

Le projet s'est amorcé au cours de l'année scolaire 2017-2018, avec pour objectif principal « *la détection d'élèves à risque de décrochage ne présentant pas de signes clairs d'enjeux liés à la persévérance scolaire* » (CSSVDC, 2022a), que le directeur général appelait affectueusement

³ Qui deviendra un centre de services scolaire en 2020.

⁴ En particulier l'associée-conseil responsable de cette restructuration et optimisation.

⁵ L'un ayant une formation de mathématicien, l'autre de statisticien. Ces deux experts quitteront Raymond Chabot Grant Thornton pour créer leur propre entreprise spécialisée en valorisation de la donnée, d'abord Stratemia, en 2020, puis Sciance, en 2022.

« *les élèves beiges* ». Les deux experts de la firme (ci-après experts externes) se sont mis à l'œuvre pour créer des modèles permettant de détecter le plus tôt possible les signes avant-coureurs du décrochage auprès d'élèves à risque. En plus d'être impressionnés par la quantité et la qualité des données, les experts externes étaient séduits par le caractère éducatif et public du projet.

Ce qui m'a beaucoup plu dans ce projet, c'est la quantité de données... C'était aussi super intéressant de faire des modèles prédictifs sur le décrochage scolaire plutôt que des modèles prédictifs... dans le privé. (Expert externe)

Pour entraîner les modèles prédictifs, une base de données (dépersonnalisées) ⁶ a été constituée à partir des dossiers des 60 000 élèves qui ont fréquenté les établissements scolaires du territoire du CSSVD, depuis 2001. Cette base de données comprenait plus de 300 variables, dont des résultats académiques ainsi que « *des statistiques relatives à l'aide financière, à l'absentéisme, aux mesures disciplinaires et aux fréquents changements d'adresse* » (Valiante, 2018). Il était désormais possible « *de croiser nos données, alors qu'avant, on les analysait en silo* » (Directeur général). Un premier cycle d'entraînement, à partir de données sur le parcours des élèves du primaire, a permis d'élaborer des modèles de prédiction sur le risque de décrochage des élèves.

Parce que les banques de données datent de 2001... on était en mesure d'avoir le parcours scolaire d'un élève, mais aussi la résultante... Ça nous permettait, aussitôt que l'élève arrivait en première secondaire, de faire une prédiction sur son potentiel de décrochage dans les 5 prochaines années... en particulier en secondaire 3, 4 et 5. (Directeur général)

⁶ « *RCGT a eu temporairement accès à une base de données contenant des renseignements personnels dépersonnalisés afin de créer l'Outil. Ces données, transférées temporairement sur son serveur à l'époque, ont été détruites depuis ... l'organisme (CSSVDC) indique qu'avant de communiquer une copie de cette banque de données à RCGT, il en a retiré 80 données considérées comme sensibles, notamment l'identité de l'enfant et de ses parents, leurs coordonnées (adresse postale et de courriel, téléphone), leurs noms d'utilisateurs pour accéder au portail et leurs alias.* ». (Commission d'accès à l'information, 2022, p. 3-4).

Les experts externes ont notamment comparé les étudiants décrocheurs aux diplômés pour repérer les signes précurseurs du décrochage. Pendant l'entraînement des modèles prédictifs, ces experts ont suivi une démarche de modélisation statistique classique.

On a suivi les étapes normales... c'est-à-dire explorer les données... sélectionner des variables... pour répondre à une question précise au départ, qui était : Peut-on détecter de façon préalable d'éventuels élèves qui vont décrocher et ne pas terminer leur secondaire ?... C'était du gros bon sens, les variables qui étaient manipulées... On a été en mesure de sortir quelques modèles prédictifs qui présentaient des bonnes performances, pour être exploitées en termes opérationnels. (Expert externe)

Quant au fonctionnement général des modèles de prédiction du décrochage scolaire que les experts externes ont conçus, le directeur général l'explique comme suit :

On prend toutes les données du primaire de l'élève et on les passe dans l'algorithme, puis ça nous donne une projection de décrochage scolaire en secondaire 2, 3, 4 et 5. Quand l'élève est rendu en secondaire 1, j'ai ma première projection en secondaire 2, 3, 4 et 5... Quand l'élève arrive l'année suivante, en secondaire 2, j'ajoute à ses résultats du primaire tous ces résultats et autres données accumulées en secondaire 1 et je les crunch à nouveau dans l'algorithme, puis ça va nous donner une prédiction pour secondaire 3, 4 et 5... plus précise... le comportement de décrochage commence à se cristalliser... La même chose après le 2e secondaire, quand l'élève arrive en 3e secondaire, on fait une projection pour 3, 4 et 5 qui est plus grande au niveau de la précision, mais on a moins de temps pour le dénouer... ainsi de suite.

Les modèles de prédiction élaborés au CSSVDC ont attiré l'attention du gouvernement. Le directeur général a d'ailleurs été invité à présenter à deux reprises (2018 et 2019) au ministère de l'Éducation le projet de modélisation des données qu'il dirige dans son centre de services scolaire. Il a aussi été invité à exposer le projet au ministre français de l'Éducation nationale de l'époque, Jean-François Blanquer, qui était en visite au Québec, en octobre 2019. Le projet du CSSVDC a également attiré l'attention des médias. Par exemple, dans un article de *La Presse*, publié le 1^{er} novembre 2018, on relate que « *l'algorithme dont s'est dotée la Commission scolaire du Val-des-Cerfs a correctement identifié 92 % des élèves qui abandonneraient en troisième*

secondaire » (Valiante, 2018). On y précise que les modèles de prédiction ont permis d'identifier « environ 90 élèves qui sont entrés au secondaire cet automne et qui présentent un risque de décrochage. Les facteurs de risque pour chaque élève ont été fournis aux écoles, afin qu'ils puissent élaborer des stratégies d'intervention » (Ibid.). Pour le directeur général, ces modèles viennent « supporter les équipes écoles et les intervenants à contrer le décrochage avant que le comportement soit cristallisé chez l'élève⁷ ».

Toutefois, l'élaboration des modèles de prédiction a posé d'importants défis au début du projet, quand il s'est agi de procéder à la collecte et au traitement des données qui ont servi à l'entraînement. En 2017, la Commission scolaire du Val-des-Cerfs n'était pas maître de ses données. Il faut savoir que la majorité des données générées par les centres de services scolaires et les commissions scolaires sont hébergées dans des systèmes informatiques appartenant à la GRICS, principal fournisseur de logiciels en éducation aux niveaux primaire et secondaire. Bien que les informaticiens du CSS « nous ont aidé à nous connecter à leurs données... on a dû faire du Feature Engineering pour essayer d'extraire manuellement l'information », en raison d'une documentation limitée de la part de la GRICS, soutient l'un des experts. Le commentaire suivant du directeur va dans le même sens :

La GRICS ne nous a pas donné le code [et la documentation] pour être capables d'aller jouer dans le cube de données. Il a fallu qu'on investisse pour faire de la rétro ingénierie et pour se servir... de données que moi, comme centre de services, je rentre dans les logiciels de la GRICS.

2. Le projet pilote : 2020 à 2021

Fort de trois années d'expérimentation et à la suggestion du ministre de l'Éducation de l'époque, Jean-François Roberge, le CSSVDC a fait une demande de financement auprès du gouvernement, en 2020, pour réaliser un projet pilote, afin d'expérimenter ses modèles de

⁷ Propos tenu lors d'une entrevue à l'émission de radio *Les grandes entrevues de Paul Arcand*, au 98,5, le 15 mars 2022.

prédiction auprès d'autres centres de services scolaires de tailles différentes et de contextes socioéconomiques variés.

Le projet pilote est... la suite logique des premiers modèles prédictifs. Ce qu'on voulait vérifier, c'est si ce type de modèle est applicable dans d'autres contextes que celui de Val- des Cerfs... par exemple, dans un milieu urbain versus un milieu rural ou dans un petit centre de services versus un grand centre de services. (Expert externe)

Le projet pilote visait donc à évaluer le potentiel d'exportabilité des modèles de prédiction utilisés au CSSVDC vers d'autres centres de services scolaires (CSS). Il visait aussi à :

1. Adapter les modèles prédictifs aux particularités d'autres CSS;
2. Évaluer le rendement des modèles;
3. Identifier de potentiels décrocheurs.

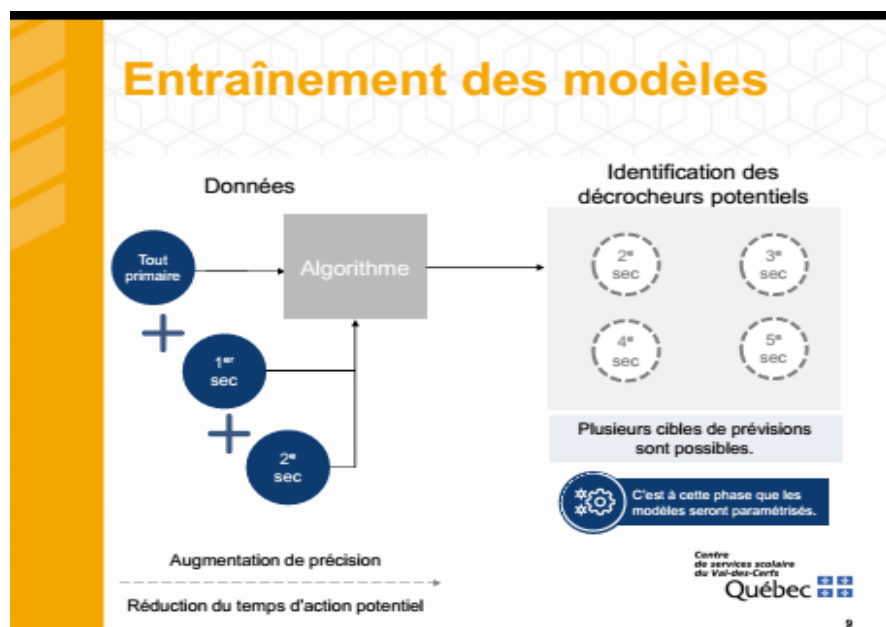
Outre le CSSVDC, cinq autres CSS ont été retenus pour le projet pilote, soit le CSS Vallée-des-Tisserands, le CSS de Saint-Hyacinthe, le CSS de Laval, le CSS de la région de Sherbrooke et le CSS des Chênes.⁸ En novembre 2020, le gouvernement s'engageait à financer le projet pilote, tout en collaborant avec les équipes du CSSVDC. « *Le ministère de l'Éducation, étant désormais associé au projet, souhaite approfondir la méthodologie, dans le but d'un jour en faire bénéficier l'ensemble du réseau scolaire du Québec.* » (CSSVDC, 2022a) À l'aide du financement gouvernemental, le CSSVDC s'est doté de sa propre infrastructure pour valoriser ses données et celles des autres CSS du projet pilote.

À partir de là, nous avons pu travailler à l'extérieur des outils de la GRICS... Dans la subvention, c'était prévu que j'achetais un serveur externe. J'ai donc monté une infrastructure, puis c'est comme ça qu'on roule en ce moment. (Directeur général)

⁸ Pour un tableau plus détaillé sur les CCS participants, voir l'annexe 1.

L'entraînement des modèles de prédiction et l'ajustement des algorithmes se sont donc

Figure 1



poursuivis, mais cette fois à plus grande échelle, avec l'ajout des cinq CSS susmentionnés. Comme dans le cas du précédent projet d'expérimentation, l'entraînement des modèles a permis d'identifier des décrocheurs selon le niveau de scolarité au secondaire (voir figure 1)⁹.

L'entraînement a aussi consisté à adapter les modèles de prédiction et leurs algorithmes au contexte d'expérimentation des 5 autres CSS participants. « On a entraîné les modèles dans chacun des centres de service de façon séparée pour qu'ils soient viables¹⁰. » (Directeur général) Quant au nombre de variables utilisées dans les modèles, il n'a pas cessé d'augmenter depuis le démarrage du projet. De 300 au début de l'expérimentation, il est passé à plus de 1500 au terme du projet pilote.¹¹ Les variables utilisées dans les modèles prédictifs ne se limitaient pas seulement au dossier de l'élève et à ses résultats scolaires gérés par le logiciel GPI de la GRICS¹². Les modèles prédictifs s'alimentaient aussi de variables non académiques gérées par d'autres logiciels de la GRICS.

On rentre aussi là-dedans les données qui viennent de Géobus¹³... qui est la distance du parcours d'autobus de l'élève... Avec Géobus, on a aussi l'adresse

⁹ Source : CSSVDC, 2020b.

¹⁰ Bien que chaque CSS participant disposait de sa propre infrastructure et de son entrepôt de données, les modèles de prédiction ont été entraînés sur le serveur du CSSVDC. « Seul VDC disposait à l'époque d'une puissance de calcul suffisante pour procéder à l'entraînement des modèles » (Expert externe)

¹¹ Au moment des entretiens, on estimait à 2000 le nombre de variables pouvant être prises en compte par les modèles.

¹² Pour plus de détails sur le logiciel GPI, consultez le lien suivant : <https://grics.ca/produit/gpi/>.

¹³ Pour plus de détails sur le logiciel Géobus, consultez le lien suivant : <https://grics.ca/produit/geobus/>.

postale. L'enfant qui a déménagé 8 fois dans son parcours scolaire, dont 4 fois dans sa dernière année, ça peut avoir un impact... On prend aussi Avant-garde¹⁴, qui est notre logiciel de service de garde pendant son primaire. Est-il allé au service de garde ? Combien de fois par semaine ? Était-il là de l'ouverture à la fermeture ?... On prend aussi des données au niveau de notre logiciel de paie. La paie est en lien avec les mesures de soutien... On considère aussi la scolarisation de la mère... Donc, ce n'est pas juste des données académiques. (Directeur général)

Pour rendre les variables plus intelligibles et plus facilement exploitables, notamment dans des

Figure 2



tableaux de bord, celles-ci ont été regroupées sous les 5 catégories suivantes : profil scolaire, comportement, assiduité, performance, sociodémographie. La figure 2 présente ces 5 catégories et leurs principales variables. « On peut forer à l'intérieur de ces catégories pour voir les

variables... qui gravitent autour de l'élève et qui ont augmenté sa possibilité de décrocher... Pour chacun de ces élèves, on est en mesure de savoir quels sont les symptômes qui ont fait que cet élève s'est retrouvé ciblé. Ce qui fait qu'on peut agir de façon plus précise sur les motifs qui pourraient l'amener vers un décrochage. » (Directeur général)

Les variables n'exercent pas toutes le même pouvoir de prédiction à l'intérieur des modèles. On estime à une trentaine le nombre de variables qui ont un impact majeur et sur lesquelles il est possible d'agir, souligne le directeur général. L'influence d'une variable peut également varier dans le temps, particulièrement quand le contexte change. C'est le cas des absences dans les écoles. Alors que les absences étaient considérées comme une variable influente dans les

¹⁴ Pour plus de détails sur le logiciel Avant-garde, consultez le lien suivant : <https://grics.ca/produit/avant-garde/>.

modèles prédictifs au début de l'expérimentation, cette variable a perdu beaucoup de son pouvoir de prédiction pendant le projet pilote (et même après), à cause du nombre élevé d'absences dues à la COVID. Pour ne pas biaiser les résultats, la variable liée aux absences (comme d'autres variables non significatives) a été supprimée des modèles de prédiction. « Une solution a été mise en place afin de filtrer les variables qui ont trop dérivé au cours du temps » (Expert interne). Pour mieux tenir compte de l'effet de la COVID sur l'influence de certaines variables dans les modèles, un outil a été créé et intégré aux modèles de prédiction. Cet outil de mitigation de l'impact de la COVID-19 a permis de gérer de manière adéquate le phénomène de dérive conceptuelle des algorithmes initiaux et d'utiliser « *les données pandémiques avec un taux de précision équivalent aux données pré-pandémiques* » (Directeur général). Avec l'ajout de cet outil, « *l'impact de la crise sanitaire peut être greffé à l'outil principal et permet de pondérer les variables pandémiques aux données d'analyses pré-pandémiques* ». (CSSVDC, 2022a) ¹⁵

Les modèles expérimentés dans le projet pilote offraient deux niveaux d'interprétation des risques de décrochage : d'une part, une interprétation globale qui inférait un niveau de risque de décrochage pour un ensemble d'élèves et, d'autre part, une interprétation individuelle qui estimait le risque de décrochage de chaque élève. Ces niveaux de risque sont estimés en exploitant l'ensemble des variables disponibles. Pour chacun des modèles retenus, les CSS ont eu aussi à choisir entre une approche prédictive dite « agressive », qui « *va cibler moins d'élèves avec un coefficient de probabilité plus haut... On est donc plus certain qu'ils vont décrocher* » et une approche dite « conservatrice », qui « *ramasse plus d'élèves dans les mailles du filet... dont des faux positifs plus nombreux... On en identifie plus, mais avec une plus grande incertitude sur le fait qu'ils soient d'éventuels décrocheurs* ». Les CSS adopteront l'une ou l'autre de ces

¹⁵ L'outil de mitigation de la dérive ne se limite pas à la COVID. « *Il élimine aussi les variables dont la sémantique évolue trop au cours du temps. L'élimination est spécifique à chaque cohorte. Les variables éliminées pour un modèle de 1^{ère} secondaire ne sont pas (nécessairement) les mêmes que pour un modèle de secondaire 2. L'outil permet de régulariser le modèle, en éliminant les variables pour lesquelles des variations mineures entraîneraient des variations dans les probabilités invraisemblablement trop fortes (cf. arbitrage biais-variance, régularisation et sensibilité).* » (Expert externe)

approches, selon les risques qu'ils sont prêts à assumer et les niveaux d'interventions qu'ils souhaitent mettre en œuvre (individualisé, sous-groupe ou général).

La phase de modélisation et d'entraînement du projet pilote a permis d'expérimenter et d'évaluer des modèles ayant un niveau de performance variable d'un CSS à l'autre, « *mais satisfaisant dans l'ensemble* » (CSSVDC et RCGT, 2021).¹⁶ Malgré certains défis d'intégration, le projet pilote semble avoir fait la démonstration qu'il était possible « *de traiter de façon opérationnelle l'ensemble des données provenant de 6 CSS dans une seule infrastructure* » et de les valoriser en tenant compte du contexte du CSS ainsi que de l'impact de la COVID. (*Ibid.*) Pour les responsables du projet, « [il] est légitime de penser que l'utilisation des techniques d'analytique avancée et de modélisation des données, telles que testées dans le projet pilote, serait possible auprès de l'ensemble des centres de services scolaires québécois ». (*Ibid.*) Pour un des experts externes, cette expansion vers d'autres CSS serait possible « *à la condition que les organisations soient responsables de la modélisation et de son intégration* » dans le fonctionnement de leur organisation. Le défi à relever dépend davantage de la maturité des organisations dans cette valorisation des données que d'enjeux strictement technologiques, précise-t-il.

3. L'après projet pilote : 2022 à 2023

Une fois le projet pilote terminé, « *certaines centres de service qui (y) avaient participé ont décidé de poursuivre l'aventure, et d'autres ont décidé de se greffer* ». (Directeur général) Le CSSVDC a embauché, en 2021, un scientifique de données¹⁷ (ci-après expert interne), pour renforcer son expertise en valorisation des données et pour faire profiter de cette expertise à d'autres CSS. Ainsi, le CSSVDC se donnait les moyens d'être plus autonome dans les suites du

¹⁶ Pour plus de détails, voir à l'annexe 2 le tableau récapitulatif des performances obtenues dans les modèles retenus par les CSS participants.

¹⁷ Ce scientifique des données occupe le poste de Directeur adjoint de service – Développement de solutions d'analytique avancée, au CSSVDC.

projet pilote. Les deux experts externes¹⁸ ont formé le scientifique de données nouvellement embauché pour qu'il soit capable d'entraîner et de faire évoluer les modèles prédictifs qu'ils avaient élaborés et de les déployer auprès des CSS intéressés à poursuivre l'aventure ainsi qu'auprès de nouveaux CSS partenaires. Une des responsabilités de cet expert interne consiste à entraîner et à évaluer les modèles de prédiction, afin de guider les CSS dans la sélection des modèles à exploiter.

À chaque rentrée scolaire, tous les modèles prédictifs (une trentaine) sont entraînés... sur un serveur qui est à Val-des-Cerfs... Je vais récupérer l'ensemble des documents que le modèle va générer et, avec ça, je vais avoir une rencontre avec les différents centres de service¹⁹, pour leur présenter les modèles... et voir avec eux lesquels sont exploitables... [pour qu'] on les rende disponibles... Si le modèle identifie seulement 30 % des élèves qui vont rater leur examen, je ne suis pas sûr que ça soit bon de le mettre en production.
(Expert interne)

Bien que les CSS aient gagné en autonomie avec l'embauche d'un scientifique de données, les deux experts externes ont continué à assumer des responsabilités dans la poursuite du projet. Une de leurs principales responsabilités a consisté (et consiste toujours) à implanter au sein des CSS l'infrastructure technologique leur permettant de valoriser leurs données. Quant au scientifique de données du CSSVDC, il assume la responsabilité de faire évoluer les modèles de prédiction et d'accompagner les équipes des CSS et des écoles dans la sélection et l'adaptation des modèles.

Au moment de entretiens, les modèles de prédiction ne se limitaient plus à la persévérance ou au décrochage scolaire. D'autres types de modèles de prédiction ont aussi été développés, dont l'un portant sur le risque d'échec des élèves à leurs examens obligatoires de 4^e ou de 6^e année du primaire en français ou en mathématiques, et l'autre, sur le risque de ne pas monter en 5^e année du secondaire et de ne pas obtenir de diplôme de fin d'études. Depuis peu, l'équipe du CSSVDC s'est engagée dans un chantier qui vise à intégrer un volet prescriptif au projet de

¹⁸ Depuis, ces deux experts ont quitté la firme Raymond Chabot et Grant Thornton, l'un en 2019 et l'autre en 2020. En 2021, ils ont créé ensemble leur entreprise, appelée « Science ». Le CSSVDC a continué à faire affaires avec ces deux experts.

¹⁹ En particulier, les directions des services éducatifs.

valorisation des données. Pour l'appuyer dans ce nouveau chantier, l'équipe du CSS du Val-des-Cerfs a fait appel au Centre de transfert pour la réussite éducative du Québec (CTREQ), qui est un organisme de liaison et de transfert en innovation sociale. Plus précisément, le CSSVDC et le CTREQ collaborent à un projet de recherche pour élaborer des modèles prescriptifs soutenant les enseignants et autres professionnels de l'éducation, dans la mise en œuvre d'interventions adaptées aux élèves à risque de décrochage scolaire. Ce volet prescriptif s'inscrit dans la continuité des modèles de prédiction élaborés jusqu'ici, soutient le directeur général :

Le projet de recherche avec le CTREQ... est pour faire le pas additionnel... Une fois qu'on a identifié un élève comme potentiel décrocheur... on veut que le coffre d'outils suive derrière pour dire : « Si tu veux soutenir l'élève, voici les modèles, selon la recherche, qui ont le plus de résultats. »

En 2022, le CSSVDC a créé un centre d'expertise régional destiné à la valorisation des données. La mission de ce centre consiste à rendre accessibles et à partager des services et des ressources en matière d'intelligence artificielle et d'affaires, pour soutenir la réussite des élèves. Le centre s'apparente à une communauté de pratique, en valorisation des données à l'échelle régionale. « *On met en commun nos ressources, on développe en commun de nouveaux algorithmes, de nouveaux tableaux de bord, puis on se les échange entre nous... On s'auto-forme entre nous.* » (Directeur général) Outre cette mise en commun, « *chaque centre de service demeure maître des modèles qu'il veut déployer* », précise le directeur général. 7 CSS avaient joint le centre d'expertise régional au moment des entretiens. De l'avis du directeur général, les 15 CSS de la Montérégie et de l'Estrie feraient à terme partie de ce centre d'expertise. Quant au fonctionnement du centre, il est encadré par une structure de gouvernance à trois paliers décisionnels :

Le CODIR, qui est le comité directeur, le COPIL, qui est le comité de pilotage, puis on a l'escouade (ou Comité de travail), qui est les petites mains sur le clavier qui réalisent le projet (COTRA). Le comité de pilotage pilote les dossiers pour être sûr que l'escouade les réalise. Puis, le comité directeur, c'est lui qui priorise les dossiers et qui donne les mandats²⁰. (Directeur général)

²⁰ Pour plus de détails sur cette structure de gouvernance, voir l'annexe 3.

Depuis un an, les travaux de valorisation des données au centre d'expertise régional s'inscrivent surtout dans une perspective d'intelligence d'affaires, avec le soutien étroit du scientifique de données. Après avoir contribué à l'élaboration de modèles de prédiction, les CSS collaborent maintenant à l'élaboration et à la mutualisation de tableaux de bord à l'intérieur de ce qu'il est convenu d'appeler un « magasin » ou un « catalogue de tableaux de bord », qui minimise considérablement la quantité de codage nécessaire²¹.

On est rendu avec un catalogue de tableaux de bord. Donc, les nouveaux CSS qui intègrent sont en mesure de choisir parmi le catalogue quels types de tableaux de bord ils veulent. Ils les sélectionnent, et leurs équipes TI prennent charge du codage et l'installent chez eux. En l'espace d'une demi-heure, ils sont capables de rouler les outils BI et de modélisation²². (Directeur général)

Le concept de « centre d'expertise » a plu au MEQ. Le ministère l'a d'ailleurs intégré à son chantier d'intelligence numérique, qui a démarré en 2021, à la différence que les centres d'expertise sont appelés « communautés de pratique ».

Le 13 mars dernier, lors d'une conférence de presse à Montréal, le ministre de l'Éducation, M. Jean-François Roberge, et son homologue responsable de la Cybersécurité et du Numérique, Éric Caire, ont valorisé et appuyé publiquement le projet (pilote). Québec est même allé plus loin, avec la volonté de numériser l'ensemble des données accumulées par les Centres de services scolaires du Québec. Un investissement avoisinant les 11 millions de dollars sur deux ans, dans le but d'appuyer les initiatives d'intelligence numérique telles que développées en Estrie par le CSS du Val-des-Cerfs et le CSS Cœur-des-Vallées (CSSVDC, 2022a)

Le MEQ entend d'ailleurs interconnecter les communautés de pratiques, pour favoriser un partage d'expertises et de ressources à l'échelle nationale.²³

²¹ Pour des exemples de tableaux de bord, voir l'annexe 4.

²² Comme le tableau de bord ne comporte « pas d'éléments spécifiques au CSS... le rajout d'une ligne (suffit) pour que son code soit installé ». (Expert externe)

²³ Pour une schématisation du réseau des centres d'expertise, voir l'annexe 5.

4. Des conditions facilitantes

Différentes conditions ont facilité le déploiement du projet, depuis l'expérimentation des premiers modèles de prédiction au CSSVDC, à la création d'un centre d'expertise régional, en passant par le projet pilote. Examinons certaines de ces conditions facilitantes.

Des données à exploiter nombreuses et variées

L'entraînement des modèles de prédiction exige l'exploitation de nombreuses données, lesquelles forment la matière première de ces modèles. Au fil du temps, la quantité de données amassées sur les élèves par le CSSVDC et par les autres CSS partenaires semble satisfaire cette exigence. C'est notamment l'avis des experts externes qui ont eu à concevoir et à entraîner les premiers modèles de prédiction. Qu'il suffise de rappeler que le CSSVDC détient des données sur les 60 000 élèves qui ont fréquenté ses établissements scolaires depuis 2001. Collectées tout au long du parcours de l'élève, ces données sont non seulement nombreuses, mais variées. Comme nous l'avons vu, les données exploitées dans les modèles ne se limitent pas aux résultats scolaires de l'élève. Elles font aussi référence à son profil scolaire, à son comportement, son assiduité et ses caractéristiques sociodémographiques. Au moment des entretiens, l'entraînement des modèles pouvait s'appuyer sur l'exploitation de 2 000 variables, et rien ne laisse croire que ce nombre ira en diminuant.

Une infrastructure technologique dédiée

Le CSSVDC s'est doté de sa propre infrastructure technologique, dont un serveur, un entrepôt de données²⁴ et des applications d'analytique avancée²⁵, ce qui a permis à cette organisation et aux autres CSS partenaires de fonctionner de façon autonome dans la valorisation des données, que ce soit pour entraîner des modèles de prédiction ou pour créer des tableaux de bord. Bien que l'extraction des données des systèmes de la GRICS ait nécessité au départ un important travail de rétro ingénierie, cette opération a valu la peine d'être menée, si l'on en juge

²⁴ *Microsoft SQL Server.*

²⁵ *Power BI.*

par l'autonomie de fonctionnement qu'elle procure au CSSVDC et à ses partenaires. La valorisation des données par les CSS peut désormais se faire localement, sans avoir à recourir à un fournisseur, comme la GRICS ou une firme externe, pour extraire les données et les exploiter. Cette infrastructure technologique procure aussi des avantages en matière de sécurité de l'information, souligne le directeur général : « *Je me suis doté de mon propre serveur pour l'expérimentation, que je débranche d'Internet pour plus de sécurité.* » (Directeur général) Dans le cas des 7 CSS du centre d'expertise régional, « *si on ne roule pas les algorithmes chez eux, on transfère leurs données qui sont encryptées (dépersonnalisées) sur notre serveur. Je le débranche du réseau, on crunch leurs données, on réencrypte (repersonnalise) les données, on rebranche, puis on retourne ça dans leur centre de service* ». (Directeur général). De plus, « *les données partagées entre les CSS sont aussi toutes pseudo-anonymisées ... Seuls les CSS peuvent dé-pseudo-anonymiser leurs données* ». (Expert externe)

L'embauche d'un expert en valorisation des données

La stratégie d'autonomisation du CSSVDC et de ses partenaires ne s'est pas limitée à l'acquisition d'une infrastructure technologique autonome. Elle s'est aussi traduite par l'acquisition d'expertises en valorisation des données, dont l'embauche d'un scientifique de données. Cette embauche a été d'autant plus bienvenue qu'une démarche de valorisation des données peut s'avérer une opération complexe au sein des organisations publiques, particulièrement en éducation. « *Ce n'est pas vrai qu'on peut juste faire un outil prédictif, déployer ça, donner ça au centre de service, puis leur dire fiez-vous à ça les yeux fermés. Ça ne marche pas comme ça ... Comme ça touche directement le parcours d'un élève, c'est sensible... C'est important que les gens comprennent les intrants... qu'il peut avoir des biais, que ce n'est pas une boule de cristal, que ce n'est pas parfait.* » (Expert externe) Le fait de disposer à l'intérieur d'un expert en valorisation des données qui connaît bien la réalité du terrain peut en effet faciliter le déploiement d'outils en intelligence artificielle (modèles de prédiction) ou en intelligence d'affaires (tableaux de bord) dans l'organisation. Le scientifique de données qui a été embauché exerce d'ailleurs différentes responsabilités auprès du CSSVDC et de ses

partenaires, comme évaluer la performance (qualité) des modèles de prédiction; accompagner les CSS dans le choix des modèles à exploiter; faire évoluer ces modèles en les adaptant aux particularités des CSS qui les utilisent; vulgariser l'interprétation des résultats de ces modèles et, dans une certaine mesure, guider les CSS dans la conception de tableaux de bord²⁶. « *Même s'il y a certains centres de service où ça fait 3 ou 4 ans qui sont dans le projet pilote, ça reste de la statistique. Tout le monde n'a pas le background de statisticien, ce qui est normal... Il faut vulgariser l'information... pour que ces personnes qui vont être porteuses dans leur centre soient vraiment à l'aise à poser leurs questions.* » (Expert interne)

La mutualisation des ressources

La mutualisation des ressources stratégiques est l'un des objectifs poursuivis par la création d'un centre d'expertise régional. La mise en partage d'un scientifique de données s'inscrit dans cet objectif. « *Ce n'est pas pensable, qu'au Québec, dans chacun des centres de service, on embauche des scientifiques de données. Même en intelligence d'affaires, vous êtes dans un contexte... où les ressources sont rares, vous devez mutualiser des ressources.* » (Expert externe) Le fait de partager le travail du scientifique de données du CSSVDC a permis à d'autres CSS de profiter d'un savoir spécialisé, sans avoir à embaucher de leur côté une telle ressource. Le CSSVDC a aussi mis à la disposition de ses partenaires son infrastructure technologique, afin qu'ils puissent valoriser leurs propres données, sans avoir à investir dans une telle infrastructure. Des efforts d'accompagnement peuvent aussi être mis en commun pour favoriser l'intégration d'un nouvel employé. « *Le CSS de la Vallée-des-Tisserands, qui est un de nos partenaires... quand il a perdu son programmeur, l'équipe autour a été capable de prendre son nouvel employé, puis de le mettre à niveau, tout en continuant de développer.* » Enfin, le fruit des opérations de valorisation des données, en particulier la conception de modèles de prédiction et la création de tableaux de bord, peut aussi être mutualisé. « *On vient de développer et de mettre en place un catalogue de tableaux de bord... Si je décide que, une journée, ça me prend un tableau de bord sur les ressources humaines... sur le plan de la main-*

²⁶ Un expert en intelligence d'affaires a aussi été embauché par le CSSVDC pour accompagner les CSS dans la conception de tableaux de bord.

d'œuvre, je vais voir dans le catalogue, et si un centre de service l'a développé, je le prends, je le descends dans mon organisation, on passe les données, et voilà, j'ai pas eu besoin de travailler. » (Directeur général)

Des modèles de prédiction à code source ouvert (*open source*)

Les modèles de prédiction dans le projet de valorisation des données ont été conçus en *open source*. Le code source de ces modèles est donc disponible pour les CSS et autres utilisateurs qui souhaitent l'examiner, l'utiliser et l'adapter. « *Je roule sur des logiciels libres... il n'y a pas de propriété intellectuelle... je n'ai rien breveté.* » (Directeur général) Ainsi, le CSSVDC et ses partenaires ont accès gratuitement aux variables et aux algorithmes des modèles de prédiction, pour les réutiliser et les faire évoluer, sans avoir à assumer des frais de licence. Dans la mesure où il s'agit « *de modèles extrêmement sensibles qui vont qualifier des élèves, c'est important qu'ils soient ouverts... que les CSS qui utilisent les modèles prédictifs... comprennent la méthodologie, comment les modèles sont faits... Ça ne peut pas être fermé, fait en arrière de portes closes.* » (Expert externe)

La présence d'un dirigeant entrepreneur

L'engagement soutenu du directeur général dans toutes les phases du projet et son leadership ont su rallier plusieurs partenaires à son projet de valorisation des données. Le directeur général a été, non seulement l'instigateur de ce projet, mais aussi son principal porteur et ambassadeur. L'importance qu'il accorde depuis longtemps à la persévérance scolaire et son expérience de décrocheur lui confèrent une crédibilité aux yeux des partenaires du projet et du MEQ. Par sa transparence, sa persévérance et sa crédibilité, il semble réunir les traits d'un leader transformationnel. Il a d'ailleurs su rallier plusieurs parties prenantes, tôt dans le projet, tout en donnant l'heure juste. « *J'ai rencontré mes 4 syndicats, puis mes 2 associations de cadres, ici, au tout début, pour leur parler du projet... en leur montrant les avantages. Mais en ne cachant pas que je pourrais avoir la possibilité de forer jusqu'à l'enseignant. Mais, c'est pas le but du projet... Je leur ai dit que je ne le ferais pas et que je voulais respecter l'autonomie des enseignants, les soutenir. Ils ont été partants.* » (Directeur général)

5. Des enjeux de valorisation

Les responsables et partenaires du projet de valorisation des données ont aussi été confrontés à des enjeux et à des défis de gestion dans la mise en œuvre de ce projet. Examinons ici quelques-uns de ces enjeux et défis.

Une culture de données et de pratiques de valorisation à instaurer

La valorisation des données n'est pas une pratique qui semble bien implantée auprès de la majorité des CSS. La vétusté des infrastructures technologiques et l'absence d'outils analytiques peuvent expliquer en partie ce retard. Pour plusieurs CSS, *« c'est extrêmement difficile à accéder à de la donnée de base, parce qu'ils n'ont pas d'outils analytiques qui leur permettent de faire ça... Il y a une dette technologique qui est accumulée au niveau de l'éducation »*. (Expert externe) Des répondants signalent aussi le manque, voire l'absence, d'une culture de données au sein de plusieurs établissements scolaires. *« La culture de données est variable et volatile d'une école à une autre. »* (Expert interne) L'instauration d'une culture de données et de pratiques de valorisation doit passer par une prise de conscience du potentiel de réutilisation des données par les gestionnaires, les enseignants et autres intervenants. Un travail de vulgarisation et de simplification des outils est également nécessaire. Par exemple, pour simplifier l'interprétation des modèles, trois seuils de risque de décrochage ont été utilisés : élevé, modéré et bas. Les pratiques de valorisation des données peuvent aussi être déstabilisantes pour les équipes informatiques. *« Les services technologiques doivent faire évoluer leurs méthodes de travail et intégrer l'IA dans les produits qu'elle développe. »* (Stratemia, 2022) Ce changement de culture et de pratiques implique l'acquisition de nouvelles compétences : *« La difficulté, ce n'est pas de faire les modèles, c'est d'amener une organisation à être capable d'utiliser ces modèles de façon adéquate... en donnant les moyens aux organisations de se monter en expertise. »* (Expert externe) Pour faciliter l'appropriation d'une culture de données, le CSSVDC et ses partenaires ont inclus dans le projet de valorisation la conception de tableaux de bord plus simples à comprendre et à exploiter que les modèles prédictifs. Toutefois, on nous met en garde contre la tentation

d'inonder les gestionnaires et autres utilisateurs de nombreux tableaux de bord. *« Il faut faire bien attention... au Death by Dashboard, c'est-à-dire d'avoir tellement de tableaux de bord, puis tellement de données, que tu noies les gens dans l'information. »* (Directeur général) On mentionne aussi le réflexe de garder pour soi les ressources issues de la valorisation des données, comme les tableaux de bord. *« La plus grosse difficulté dans le catalogue [de tableaux de bord], c'est de faire participer les centres de service dans une mise en commun »,* tout en leur demandant de respecter une certaine méthodologie. (Directeur général)

Des préoccupations professionnelles et éthiques

La valorisation des données soulève des préoccupations de nature professionnelle, notamment auprès des enseignants. Qu'il s'agisse de modèles de prédiction ou de tableaux de bord, les enseignants peuvent percevoir ces outils comme des moyens mis à la disposition de la direction pour exercer une surveillance étroite de leur travail. Ils peuvent aussi considérer les modèles de prédiction comme une menace à leur autonomie et à leur jugement professionnel. Il n'est pas surprenant que les enseignants forment un des groupes les moins favorables au déploiement de ces outils. Au moment des entretiens, les enseignants n'avaient toujours pas accès aux modèles de prédiction. *« Depuis environ 2 ans, les directions d'école les voient (les modèles de prédiction), mais on n'était pas descendu aux enseignants encore. Depuis un an et demi, les services éducatifs les utilisent. »* (Directeur général) La valorisation des données soulève aussi des questions d'ordre éthique, en particulier par rapport aux principaux concernés, à savoir les élèves eux-mêmes. À ce propos, il peut être justifié de soulever les questions suivantes : L'utilisation de modèles de prédiction peut-elle accentuer le phénomène d'étiquetage et de stigmatisation d'élèves déjà aux prises avec des difficultés d'ordre scolaire ? Comment s'assurer que les modèles et leurs algorithmes ne reproduisent pas (ou le moins possible) des biais cognitifs et sociaux ? Les parents des enfants ciblés par les modèles doivent-ils être mis au courant ? Si oui, comment ? Doit-on demander le consentement des parents ? Outre les enseignants, qui aura accès à ces renseignements sensibles ? Les utilisateurs seront-ils en mesure d'interpréter les résultats des modèles et de savoir quand les modèles se trompent ? Voilà autant de questions auxquelles les responsables du projet de

valorisation des données doivent apporter des éléments de réponses, notamment dans le cadre d'une gestion individualisée des élèves à risque.

Des données à intégrer et à normaliser

Si les CSS utilisent les mêmes logiciels de la GRICS, il est possible de personnaliser certaines valeurs saisies localement. Toutefois, cette personnalisation soulève des enjeux d'ordre sémantique, dans la mesure où la définition des variables utilisées dans les modèles prédictifs peut changer d'un CSS à l'autre. C'est le cas du concept de décrochage : « *On s'est rendu compte qu'il y avait des définitions qui étaient différentes entre les CSS dans la conception du décrochage.* » (Directeur général) C'est aussi le cas des motifs d'absence. Par exemple, dans le système GDI de la GRICS, « *on prévoit 15 motifs d'absence différents. Quand j'ai commencé à expérimenter, en 2017, j'avais 115 descriptions d'absence différentes dans mon CSS, et les autres CSS avaient sensiblement le même nombre... Il a fallu s'entendre sur un regroupement, diminuer ce nombre de variables.* » (Directeur général) Des défis de normalisation et d'interopérabilité des données se posent, notamment quand un CSS intègre le centre d'expertise régional. Pour que les variables et leurs valeurs soient exploitables, et « *que les algorithmes... soient comparables d'un centre de service à l'autre, il a fallu se donner des définitions similaires, puis réduire le nombre de motifs... Donc, il y a eu un nettoyage de données... Parce qu'il faut parler le même langage.* » (Directeur général) Bien que des définitions différentes puissent subsister pour refléter les règles d'affaires propres à chaque CSS, « *une harmonisation doit être faite* ». (Expert interne)

Les exigences de la Commission d'accès à l'information

La Commission d'accès à l'information (CAI) a démarré une enquête de sa propre initiative sur le CSSVDC, en février 2019. « *Cette enquête fait suite à la parution d'un article de journal²⁷ affirmant que l'organisme (CSSVDC) a mis au point un algorithme permettant de cibler les élèves de 5^e année qui présentent un risque important de décrochage ... en collaboration l'entreprise Raymond Chabot Grant Thornton.* » (Commission d'accès à l'information, 2022, p. 1)

²⁷ VALIANTE, G., « Un algorithme repère les décrocheurs », *La Presse*, 1^{er} novembre 2018.

De l'avis du directeur général, « *les plus grands freins sont venus de la Commission d'accès à l'information... À peu près à tous les 6 mois, elle m'envoyait 2 pages de questions... (auxquelles) je devais répondre... sur plusieurs pages* ». Au terme de son enquête, la Commission a émis un avis d'intention, le 5 novembre 2020, dans lequel elle indique que, même si plusieurs mesures de protection des renseignements personnels ont été prises durant le développement des modèles prédictifs, le CSSCVD n'a pas respecté « *certaines exigences de la Loi sur l'accès, plus précisément en ne rencontrant pas son obligation d'information auprès des parents et en n'ayant pas établi de délai de destruction spécifique pour les résultats produits dans le cadre du projet* » (*Ibid.*, p. 2). Dans une décision émise le 9 novembre 2022, la Commission ordonne au CSSVDC « *d'informer les parents d'élèves dont les renseignements personnels dépersonnalisés ont servi au développement de l'Outil* » (*Ibid.*, p. 13). De plus, elle invite le CSSVDC à « *détruire le fichier Excel (de données dépersonnalisées sur les élèves) produit dans le cadre du développement de l'Outil* » (*Ibid.*, p. 13). Enfin, elle recommande au CSSCVD de réaliser une évaluation des facteurs relatifs à la vie privée, avant de procéder au déploiement de l'outil.

Conclusion

La valorisation des données en éducation, en particulier la conception et l'utilisation de modèles de prédiction, semble un moyen efficace de repérer et de gérer les problèmes qui peuvent survenir dans le parcours scolaire des élèves. La performance de ces outils d'intelligence artificielle dépend de la qualité et de la quantité des données qui servent à leur entraînement. Les CSS détiennent suffisamment de données de qualité pour entraîner les modèles de prédiction et alimenter d'autres outils dits « intelligents », comme les tableaux de bord. Mais encore faut-il que ces données soient facilement accessibles aux CSS, et que ceux-ci aient les capacités technologiques et humaines de les exploiter. En se dotant d'une infrastructure technologique et en embauchant un scientifique de données, le CSSVDC s'est donné les moyens d'extraire et d'exploiter de manière autonome ses propres données. Il n'en demeure pas moins que la valorisation des données suppose une culture, des compétences et des pratiques encore peu présentes en éducation, comme dans bien d'autres secteurs. Il ne suffit pas de concevoir et de rendre disponibles des modèles de prédiction pour que les CSS et leurs établissements se les approprient. S'approprier des modèles de prédiction demande une maturité organisationnelle qui doit s'acquérir. Pour faciliter l'acquisition d'une culture et d'une expertise en valorisation des données, certains proposent de commencer par le déploiement de tableaux de bord, comme étape transitoire vers une appropriation des modèles de prédiction. *« Une bonne passerelle vers les modèles prédictif, c'est de proposer les tableaux de bord qui... sont plus facilement compréhensibles par les directions... et les directions adjointes. »* (Expert externe) Quant au déploiement de modèles de prédiction, des répondants suggèrent d'avancer avec prudence : *« C'est quand même des prédictions qui touchent des données relativement sensibles. On veut que les gens soient dubitatifs. On veut que les prédictions soient validées. »* (Expert externe) Bien que les modèles de prédiction puissent générer une probabilité de décrochage pour chaque élève, il peut être prudent, pour l'instant, de gérer de façon agrégée les risques de décrochage. *« Sans nécessairement descendre au niveau de l'élève, les prédictions agrégées restent pertinentes et permettent quand même la mise en place d'interventions par le centre de service. »* (Expert externe) Si les outils

d'intelligence artificielle peuvent aider à la détection des cas à risque de décrochage en éducation, ils doivent être subordonnés au jugement des intervenants sur le terrain, soutiennent les répondants. Autrement dit, il importe d'utiliser les modèles de prédiction comme des outils d'aide à la décision et non comme des mécanismes de surveillance et de sanction. « *Le but, ce n'est pas de trouver le coupable, le but c'est de soutenir l'élève dans son cheminement scolaire... On a développé un outil d'aide à la décision. On n'a pas développé un outil qui décide pour nous.* » (Directeur général) L'adoption d'une approche non coercitive a d'ailleurs permis au directeur général d'obtenir l'adhésion de plusieurs parties prenantes au sein de son organisation. Le récent volet prescriptif du projet, qui mise sur des interventions scolaires adaptées, plutôt que sur l'évaluation du rendement des enseignants, tend à confirmer une telle approche. Au-delà de l'adhésion des parties prenantes à l'interne, le CSSVDC devra aussi convaincre et rassurer les parents que ces modèles de prédiction se révéleront utiles à la réussite scolaire de leur enfant, tout en minimisant les effets d'une possible stigmatisation et autres préjudices.

Bibliographie

Centre de services scolaires du Val-des-Cerfs (2022a). *L'intelligence artificielle au service de la persévérance scolaire*, Communiqué publié le 18 mars 2022, [en ligne] <https://cssvdc.gouv.qc.ca/lintelligence-artificielle-au-service-de-la-pseverance-scolaire/>

Centre de services scolaires du Val-des-Cerfs (2022b). *Analytique avancée et intelligence artificielle en éducation*, Présentation de l'initiative du CSS du Val-des-Cerfs publiée le 9 juin 2022.

Centre de services scolaires du Val-des-Cerfs et Raymond Chabot et Grant Thorton (CSSCDV et RCGT) (2021), *Analytique avancée et intelligence artificielle en éducation*, document (diaporama) présenté le 15 octobre 2021 à la Table de concertation régionale de la Montérégie.

Commission d'accès à l'information (2022). « Décision » émise le 9 novembre 2022 à la suite de l'enquête en vertu de la *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels*.

Stratémia (2022). « Enjeux éthiques dans le cadre d'utilisation de l'intelligence artificielle au sein des organisations scolaires »

Valiante, Giuseppe (2019). « Un algorithme repère les décrocheurs », publié dans *La Presse*, le 1^{er} novembre 2018, [en ligne], <https://www.lapresse.ca/actualites/education/201811/01/01-5202518-un-algorithme-repere-les-decrocheurs.php>

Annexes

Annexe 1

CCS partenaires dans le projet pilote

		CSSSH	CSSRS	CSSDC	CSSVT	CSSDL	CSSVDC
Géographie		MRC d'Acton et des Maskoutains	Région de Sherbrooke	MRC de Drummond	M.R.C. de Beauharnois-Salaberry et Le Haut-Saint-Laurent.	Ville de Laval	MRC de Brome-Missisquoi et de la Haute-Yamaska
Nombre employés		3000	3790	2 300	1 800	10 000	3 000
Nombre élèves		17 000	25 000	15 000	9 000	56 000	19 350
Nombre établissements		38	47	55	34	85	43
Préscolaire et primaire	Élèves	-	13 207	7 862	5 681	28 484	10 350
	Établissements	30	38	35	27	58	33
Secondaire	Élèves	-	6 715	1 451	3 031	15 140	6 000
	Établissements	4	4	5	4	14	7
Formation professionnelle	Élèves	-	2 100	1054	674	6 711	3 000
	Établissements	1	1	1	1	8	2*
Éducation générale aux adultes	Élèves	-	3 189	1 256	375	5 852	*
	Établissements	1	1	1	1	4	*
Autres établissements		<ul style="list-style-type: none"> * 1 école spécialisée * 1 école stages-études 	<ul style="list-style-type: none"> * 1 école pour rattracheurs * 2 écoles spécialisées 		<ul style="list-style-type: none"> * 1 école primaire et secondaire 	<ul style="list-style-type: none"> * 1 service aux entreprises et de formation continue Kopus 	<ul style="list-style-type: none"> * Centre régional intégré de formation / Éducation des adultes et formation professionnelle

Annexe 2

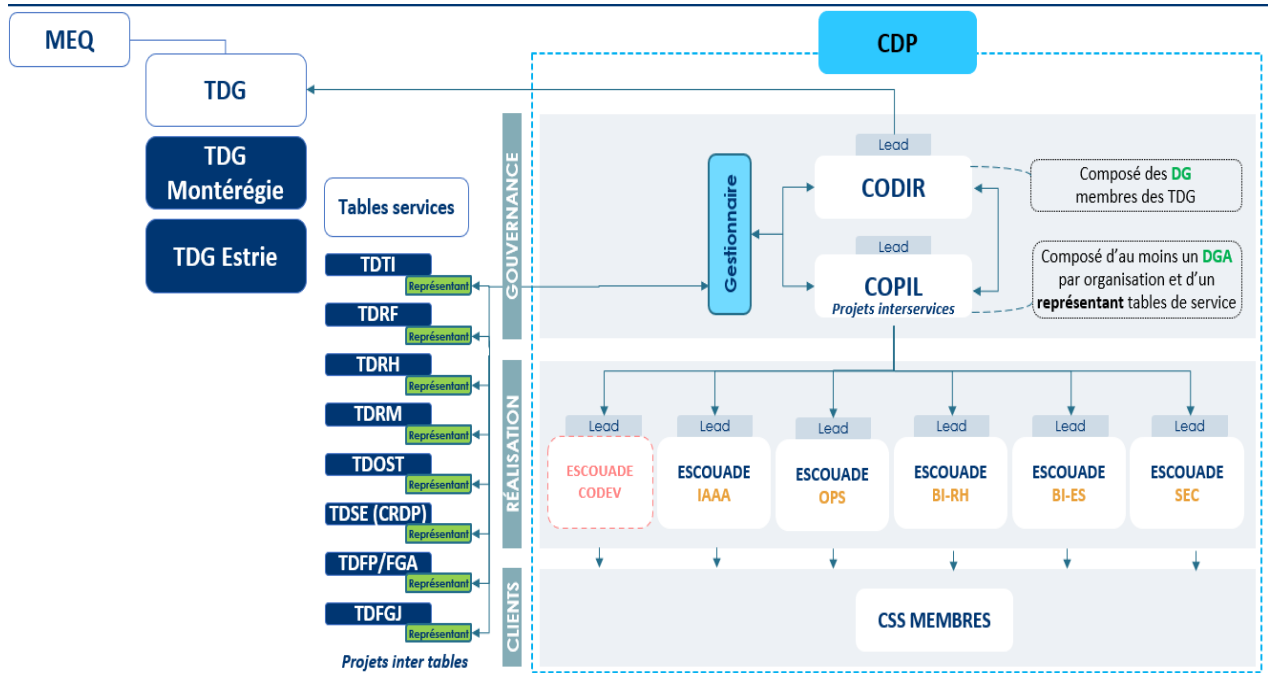
Performances obtenues pour les modèles retenus par les CSS du projet pilote

	Modèles produits	Types de modèles	Niveau d'exploitation : Adéquat / bon / excellent	Nombre total de modèles
CSSSH	p6_s2_to_s4 s1_s2_to_s5 s2_s3_to_s5	XGBOOST LASSO LASSO	Adéquat Adéquat Adéquat	7
CSSRS	p6_s2_to_s5 s1_s2_to_s5 s2_s3_to_s5	LASSO sur variables présélectionnés XGBOOST XGBOOST	Adéquat Adéquat Excellent	10
CSSVT	p6_s2_to_s5 s1_s2_to_s5 s2_s3_to_s5	XGBOOST LASSO XGBOOST	Excellent Excellent Adéquat	12
CSSDC	p6_s2_to_s5 s1_s2_to_s4 s2_s3_to_s5	LASSO sur variables présélectionnées LASSO XGBOOST	Adéquat Bon Excellent	9
CSSDL	s1_s3 s2_s3	LASSO XGBOOST	Adéquat Adéquat	4

CSSVDC	p6_s2_to_s5	<i>XGBOOST</i>	Adéquat	4
	s1_s2_to_s5	<i>LASSO</i>	Bon	5
	s2_s3_to_s5	<i>LASSO</i>	Bon	4
	s4_s5	<i>LASSO</i>	Bon	3

Annexe 3

Structure de gouvernance



Modèle de gestion

Soutenir la création de centres d'expertise



Source : Centre de services scolaires du Val-des-Cerfs (2022b)

enap.ca

QUÉBEC

MONTRÉAL

OUTAOUAIS

SAGUENAY

TROIS-RIVIÈRES

