English version below.

Fiche projet en R & D
Sciences des données - IA

Renseignements généraux

OFFRE:

- Étudiant au doctorat ou à la maîtrise
- 4 mois temps partiel ou temps plein (Environ 225 h)
- 10 000$ en bourse

Été 2023

Nom et prénom du responsable en entreprise : SABOUNJI Omar

Courriel du responsable en entreprise : omarsabounji3@gmail.com + en CC. entrepreneuriat@ivado.ca

Titre du stage : Conception de métriques linguistiques pour le suivi de la progression dans le développement des capacités rédactionnelles des apprenants dans le milieu scolaire

Mots clés:

- Linguistique
- Analytique
- NLP
- Design de métrique

Description de l'entreprise:

L'objectif global de cette start-up est de concevoir une plateforme edtech visant à personnaliser l'apprentissage pour aider les jeunes apprenants en difficulté académique. Pour ce faire, notre proposition de valeur est de contribuer à faire changer de paradigme en éducation en faisant du succès académique des apprenants, une question de progression personnelle en quantifiant granulairement chaque progrès quelle que soit sa dimension. Pour son premier produit, la startup a pour objectif d'aider à améliorer les compétences liées à la rédaction pour les jeunes apprenants du primaire au secondaire.

Description du projet:

Le concept de produit de la startup est bâti sur trois composantes : une composante de collecte de données, une composante de traitement de données et une composante de recommandation de contenu. La composante de collecte de données vise à permettre de collecter de différentes sources liées aux productions écrites des jeunes apprenants pour permettre la correction par l'enseignant.

Le projet de recherche portera sur la composante de traitement de données. Cette composante est axée sur de l'analytique et de la data science. Le développement des capacités rédactionnelles des jeunes apprenants est un construit qui se développe tout au long de l'année scolaire et d'année en année, tout au long de la scolarité jusqu'à la fin du secondaire. En se basant sur la vision que cet apprentissage est un construit, il est alors possible de savoir précisément et en temps réel, en utilisant de l'analytique, où en est l'apprenant dans le développement de ses capacités rédactionnelles. Actuellement, dans les écoles, ces capacités rédactionnelles sont évaluées de différentes manières dépendamment du niveau scolaire, de l'enseignant, de l'école et du programme scolaire. Cela rend souvent très difficile l'utilisation des données d'évaluations actuelles pour quantifier, de manière consistante et de manière très précise, les progrès des apprenants dans le développement de leurs capacités rédactionnelles. Ce constat est lié aux nombreux biais que le mode d'évaluation actuel peut générer ainsi que de la qualité et de la précision de ce mode d'évaluation. Actuellement, les barèmes de corrections sont composés de parties objectives comme les conventions linguistiques (rule-based evaluation) et des parties plus subjectives comme l'organisation ou la cohérence des productions écrites pour lesquelles l'IA et le NLP permettent de quantifier de plus en plus le niveau de progression.

L'objectif de ce projet de recherche est de construire une large gamme de métriques diversifiée pour intégrer le plus possible tous les aspects objectifs ou subjectifs qui mesurent la progression des apprenants dans leurs développements de leurs capacités rédactionnelles. On ambitionne d'être capable de les quantifier granulairement et

de façon consistante. Par exemple, pour l'évaluation de la maitrise des conventions linguistiques, une métrique très simple ressemblerait à un taux d'erreur sur 100 mots pour mesurer la maitrise d'un aspect spécifique des conventions. Pour schématiser l'objectif ultime, cette large gamme de métrique complète permettrait ultimement de suivre précisément la progression d'un apprenant tout au long de sa scolarité. On souhaite aussi s'inspirer de l'analytique utilisé pour suivre la performance dans le sport de haut niveau ou dans le jeu vidéo pour designer de tel métrique pour suivre la progression des apprenants. Cette composante analytique doit évidemment englober le plus possible les critères d'évaluations actuellement présents dans les barèmes d'évaluations fournis par les ministères de l'Éducation au Canada ou encore par le CCSS(The Common Core State Standards) aux États-Unis.

Les compétences recherchés:

- Traitement naturel du langage (NLP)
- Statistiques et analyse de données
- Visualisation de données

R & D Project Sheet
Data Science - AI

Informations

OFFER:

- Ph.D. or Master's student.
- 4 months part time or full time (Approximately 225 h)
- 10 000$ in scholarship

Summer 2023

Person in charge at the company: SABOUNJI Omar

Emails: omarsabounji3@gmail.com + en CC. entrepreneuriat@ivado.ca

Title of the project: Design of language metrics for monitoring progress in the development of learners' writing skills in the school setting

Key words:

- Linguistics
- Analytics
- NLP
- Metric design

Company description:

The overall objective of this start-up is to design an edtech platform to personalize learning to help young learners in academic difficulty. To do so, our value proposition is to contribute to a paradigm shift in education by making learners' academic success a matter of personal progress by granularly quantifying each progress regardless of its dimension. For its first product, the startup aims to help improve writing skills for young learners from elementary to high school.

Project description :

The startup's product concept is built on three components: a data collection component, a data processing component and a content recommendation component. The data collection component aims to collect different sources related to the written productions of young learners to allow correction by the teacher.

The research project will focus on the data processing component. This component focuses on analytics and data science. The development of young learners' writing skills is a construct that develops throughout the school year and from year to year, all the way through high school. Based on the view that this learning is a construct, it is then possible to know precisely and in real time, using analytics, where the learner is in the development of his/her writing skills. Currently, in schools, these writing skills are assessed in different ways depending on the grade level, the teacher, the school and the curriculum. This often makes it very difficult to use current assessment data to consistently and accurately quantify learners' progress in developing their writing skills. This is related to the many biases that the current mode of assessment can generate as well as the quality and accuracy of that mode of assessment. Currently, the evaluation scales are composed of objective parts such as linguistic conventions (rule-based evaluation) and more subjective parts such as the organization or the coherence of the written productions for which AI and NLP allow to quantify more and more the level of progress.

The objective of this research project is to build a wide range of diversified metrics to integrate as much as possible all the objective or subjective aspects that measure the progress of learners in their writing skills development. The ambition is to be able to quantify them granularly and consistently. For example, for the evaluation of the mastery of linguistic conventions, a very simple metric would look like an error rate on 100 words to measure the mastery of a specific aspect of the conventions. To summarize the ultimate goal, this broad range of comprehensive metrics would ultimately allow for accurate tracking of a learner's progress throughout their schooling. We would also like to draw inspiration from the analytics used to track performance in high-level sports or in video games to design such a metric to track learners' progress.This analytical component must obviously encompass as much as possible the evaluation criteria currently present in the evaluation scales provided by the ministries of education in Canada or by the CCSS (The Common Core State Standards) in the United States.

Skills :

- Natural Language Processing (NLP)
- Statistics and data analysis
- Data visualization