**Proposition d’Idée de Projet**

**1. Idée du Projet**

**Titre** : *" Prédire la réussite académique à partir des habitudes quotidiennes : une approche par machine learning "*

**Problème ciblé** : De nombreux étudiants présentent des résultats académiques faibles sans que la cause soit uniquement académique. Ce projet vise à identifier les habitudes quotidiennes (sommeil, études, réseaux sociaux, alimentation, santé mentale...) qui influencent significativement les résultats scolaires.

**Objectif** : Construire un modèle prédictif basé sur des données réelles simulées pour estimer la performance académique des étudiants (score d’examen) à partir de leurs habitudes de vie.

**2. Pertinence par rapport aux Objectifs de Développement Durable (ODD)**

Ce projet contribue aux ODD suivants :

* **ODD 3 – Bonne santé et bien-être** : Met en évidence les liens entre sommeil, santé mentale et performance.
* **ODD 4 – Éducation de qualité** : Identifie les facteurs extra-académiques affectant la réussite scolaire, ce qui peut informer les politiques éducatives.
* **ODD 10 – Réduction des inégalités** : Permet de comprendre comment les conditions de vie (ex. : emploi à temps partiel, qualité d’internet) affectent les étudiants de manière différenciée.

**3. Exemples de Littérature**

1. **"Impact of lifestyle factors on student academic performance"** – étude analysant comment le sommeil, l’exercice physique et l’alimentation affectent les résultats scolaires (Elsevier, 2020).
2. **"Machine learning models for predicting student performance"** – article sur l’utilisation de la régression linéaire, forêts aléatoires et réseaux de neurones pour prédire les scores des étudiants (IEEE Access, 2021).

Ces études inspirent notre approche tant sur le plan des variables à considérer que sur les méthodes de modélisation.

**4. Description de Vos Données**

* **Source**: Dataset Kaggle – *Student Habits vs Academic Performance*
* **Format** : CSV
* **Taille** : 1 000 observations, 16 variables
* **Variables** : Heures d’étude, sommeil, réseaux sociaux, alimentation, exercice, santé mentale, qualité d’internet, score d’examen, etc.
* **Prétraitement requis** : Encodage des variables catégorielles, normalisation, gestion des valeurs manquantes.

**5. Approche (Apprentissage Machine ou Deep Learning)**

* **Approche choisie** : **Apprentissage automatique (machine learning)**, notamment régression linéaire, régression Ridge/Lasso, forêt aléatoire, XGBoost.
* **Justification** : Les données sont tabulaires, peu volumineuses et bien structurées. Une approche d’apprentissage automatique classique est plus efficace et interprétable que les méthodes de deep learning dans ce contexte.