# Envision Datathon: ROUND-1

# Student Performance Predictor:

## Problem Statement

The objective of the datathon is to develop a Student Performance Predictor that can accurately forecast students' CGPA based on their academic performance, daily routines, and other relevant factors. This project aims to assist educators and institutions in identifying patterns and supporting students in achieving academic success.

## Round 1: Data Collection

In this round, we have to gather and prepare data to train the machine learning model effectively. The data collection process involves:

## 1. Survey Data

- A detailed survey was designed to collect data on students' academic records, study habits, daily routines, and other relevant attributes.  
- The survey ensured diversity in responses by targeting students from various backgrounds, age groups, and academic disciplines.

### 2. Mockaroo Data Generation

Synthetic data was generated using Mockaroo, a platform that allows the creation of realistic datasets for testing and training purposes, to augment the dataset.  
- This approach provided flexibility in tailoring datasets to specific attributes required for the model.

### 3. Additional Dataset Resources

For further enrichment of the dataset, the following resources (other than Kaggle) can be utilized:

- UCI Machine Learning Repository: https://archive.ics.uci.edu/ml/index.php

- Data.gov: https://www.data.gov/

- Open Data Portal: https://www.opendataportal.org/

- Google Dataset Search: https://datasetsearch.research.google.com/

- UNESCO Institute for Statistics (UIS): <http://uis.unesco.org/>

- government student portal: https://www.data.gov.in/sector/Education

## Goals for Round 1

- Ensure data diversity and relevance by combining real-world survey responses with synthetic data.  
- Preprocess the data to ensure consistency, remove missing values, and standardize formats.  
- Validate the data quality to prepare for feature engineering and model training in subsequent rounds.

## Tools and Technologies

- Survey Platforms: Google Forms, Typeform  
- Synthetic Data Generator: Mockaroo  
- Data Processing Tools: Python (pandas, numpy), Excel  
- Visualization Tools: matplotlib, seaborn

## Expected Outcomes

By the end of Round 1, I will have a well-structured, high-quality dataset ready for training the machine learning model, laying a strong foundation for accurate predictions in subsequent rounds.