**IST 707 Project Proposal**

Group Members: Emily Brich, Sydney Rickard, Kristan Rodriguez, Jason Rhodeman

#### **Dataset:**

Big\_student\_clear\_third\_version.csv

This dataset was found on Kaggle and includes 416,921 observations of 22 variables of MOOC (massive open online courses) data. Online learning has exploded in the last decade, and millions of people have enrolled in these courses, but how many stay with their programs and courses until the end? With this dataset, we can take a look at dropout rates and explore the different factors associated with high dropout rates.

#### **Purpose:**

The project aims to explore dropout rates in online learning, specifically in MOOCs offered by HarvardX and MitX. With the rapid growth of online education, understanding the factors influencing student persistence and dropout rates is crucial. This analysis seeks to identify patterns and associations that could inform strategies to enhance student engagement and course completion rates.

#### **Data Dictionary:**

* Institute - Where are they learning? In this case, either HarvardX or MitX
* Course\_id - Course ID
* Year - year of the course
* Semester - semester of the course
* userid\_DI - User ID
* Viewed - watched more than half of the videos
* Explored - Learned more than half of the chapters
* Certified - Whether it is qualified
* final\_cc\_cname\_DI - Country or region
* LoE\_DI - Level of education
* Gender - Gender of learner
* Grade - Test score
* start\_time\_DI - course start time
* last\_event\_DI - course end time
* Nevents - login learning times
* Ndays\_act - active times
* Nplay\_video - number of videos watched
* Nchapters - number of learned chapters
* Nforum\_posts - number of posts
* Incomplete\_flag - incomplete sign
* Age - ager of learner

#### **Analytical Techniques:**

* **Clustering:** To identify groups of learners with similar characteristics or behavior patterns.
* **Association Rule Mining:** To discover interesting relationships between variables in the dataset.
* **Decision Tree Analysis:** To model decisions and predict dropout rates based on various factors.

#### **Objectives:**

* To understand the factors contributing to high dropout rates in MOOCs.
* To identify patterns and trends in learner engagement and performance.
* To provide insights for course providers to improve course design and learner support mechanisms.

#### **Expected Outcomes:**

* A comprehensive analysis of dropout rates in MOOCs.
* Identification of key factors influencing course completion.
* Recommendations for enhancing learner engagement and course effectiveness.