



**UNB**

**University of New Brunswick**

CS4403

TERM PROPOSAL

INSTRUCTOR: JAKE VAN DER LAAN

*Author*

**Yasin Rahman**

## **1.Data Set Derivation**

I plan to use a global education data collection that covers 202 different locations and nations worldwide. The data set was obtained from Kaggle.

Link: [Kaggle.com](https://www.kaggle.com)

## **2. Dataset Size and Mining Feasibility**

The dataset used in this study consists of 202 entries, each representing a country or region, with 29 attributes related to education, literacy, and socio-economic indicators. Given the variety of features available, this dataset is sufficiently large and diverse to enable effective data mining. The number of observations allows for meaningful statistical analysis and model building without excessive risk of overfitting.

## **3. Information to Mine and Proposed Use**

This the study's main goal is to uncover information on educational attainment and the socioeconomic effects of it. This study specifically intends to investigate the relationships between job levels, birth rates, and gender differences in education and variables including literacy rates, school enrollment, and completion rates. It is suggested that this data be used to educate educational development groups, educators, and legislators about important patterns and possible areas of action.

## **4. Usefulness and Expected Insights**

This dataset provides significant information in a number of ways. First, a better understanding of the connection between education and work. Furthermore, examining gender differences might assist in pinpointing areas that need focused actions to reduce the educational gap. Finally, examining the relationship between

birth rates and educational attainment rates might help with long-term strategic planning for educational facilities.

## **5. Timeline**

**February 11, 2025** : Proposal Submission

**February 18 - February 20** : Data collection, cleaning, and preprocessing

**February 19 - February 25** : Exploratory Data Analysis (EDA) and feature selection

**February 26 - March 3** : Initial model selection and development

**March 5 - March 15** : Refinement and optimization of models

**March 18 - March 24** : Interpretation of results and drafting the final report

**March 25 - March 31** : Final adjustments and documentation

**April 1 - April 7** : Final testing

**April 8, 2025** : Project submission