**PHASE 1**

**Project to Track, Examine, and Predict School Dropouts: “A Major Concern in Canada”**

**Introduction**

According to the news article from ctvnews 2 in 5 students are seriously considering dropping out.This worrying figure reveals a rising tendency among young people in Canada, where a variety of variables influence their choice.

Students may begin to doubt the purpose and worth of their education due to the quickly changing work market and the mismatch between education and employment chances.

The decision to drop out from school can have a profound impact on individuals, this can limit their potential for future earnings, diminish their chances for a successful career, and raise their chance of falling into poverty, being unemployed, and being alone in society. Early dropout can also have a detrimental impact on one's well being and mental health.

**Problem Statement**

School dropout rates in Canada, a developed nation, remain a pressing concern. Numerous factors contribute to this issue, and our study aims to investigate these underlying causes. By analyzing various influential factors, we seek to develop a machine learning model capable of predicting student dropout. This predictive capability will equip educational authorities with valuable insights, enabling them to implement **targeted interventions and address the root causes of dropout,** ultimately improving educational outcomes and fostering a **more equitable society.**

**Business Goals**

The business goals we aim to accomplish through this project are:

* Enhance Allocation of Resources based on **anticipated dropout rates,** provide resources to support services in an efficient manner.
* To analyze various **factors responsible for drop out of students in and after post-secondary education in Canada** and also to improve Student Retention i.e.,Identify at-risk students early to provide targeted support and prevent dropouts.
* To predict or forecast the students who are likely to drop out based on these factors.
* To compare the accuracy of various supervised machine learning algorithms for predicting drop out students or classifying the students as drop outs or not dropouts.

Research Questions:

* What factors contribute to student dropout, and how do they impact it?
* Which students are most in need of support to stay in school?
* What are the core factors responsible for drop out of students in and after post-secondary education in Canada ?

**Data dictionary of selected data explaining all the attributes**

* <https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention?resource=download>
* <https://www150.statcan.gc.ca/n1/pub/81-582-x/81-582-x2024001-eng.htm>
* [Over 1-in-4 (26%) Canadian post-secondary students said they have considered dropping out of their program because of money, new poll finds (newswire.ca)](https://www.newswire.ca/news-releases/over-1-in-4-26-canadian-post-secondary-students-said-they-have-considered-dropping-out-of-their-program-because-of-money-new-poll-finds-831703839.html)

Note :We are going to build a data set from the above three sources

**List of independent features and name of dependent attribute**

Dependent Variable

|  |  |  |
| --- | --- | --- |
| Variable | Type | Description |
| Dropout Status | Categorical | Whether the student dropped out (Yes/No) |

Independent Variables

Numerical Independent Variables

|  |  |
| --- | --- |
| Variable | Description |
| Age at enrollment | The age of the student at the time of enrollment in postsecondary education. |
| High school GPA | Grade point average from high school, indicating academic performance. |
| First-year GPA | Academic performance in the first year of postsecondary education. |
| Distance from home province | Distance from home province/territory for students studying out of province. |
| Transfer credits | Number of transfer credits received from other institutions. |
| Number of dependents | The number of dependents the student has. |

Categorical Independent Variables

|  |  |
| --- | --- |
| Variable | Description |
| Indigenous status | Whether the student identifies as First Nations, Métis, or Inuit. |
| Language proficiency | Proficiency in English or French, relevant to Canada's bilingual context. |
| Academic goals | Students' educational aspirations (e.g., university, college, trade school). |
| Financial aid status | Type of financial assistance received (scholarships, loans, social assistance). |
| Parental support | Whether the student receives financial support from parents. |
| Marital status | The marital status of the student (single, married, divorced). |
| Type of postsecondary program | Type of program enrolled in (university, college, trade school). |
| Campus involvement | Participation in volunteering or social activities on campus. |
| Support network | Whether the student has someone on campus to discuss personal issues. |
| Prior dropout history | Any history of dropping out in high school. |
| Substance use history | Past drug abuse or substance use issues, particularly in high school. |
| Part-time vs. full-time status | Whether the student is studying part-time or full-time. |
| First-generation student status | Whether the student is the first in their family to attend postsecondary education. |
| Socioeconomic background | Categorized as low, middle, or high socioeconomic status. |
| Military status | Whether the student is a current or former military member. |
| Online vs. in-person learning | The mode of course delivery (online or in-person). |
| Participation in pre-university programs | Involvement in college preparation courses before enrollment. |
| Institutional commitment | Student's level of commitment to their chosen institution. |
| Work status during studies | Whether and how much the student works while studying (full-time/part-time). |

**Methodology :**

For this project we will be using **Supervised Classification** as it is the most suitable **machine learning technique for predicting post-secondary dropouts in Canada**. By using a supervised classification algorithm, we can train the model on a dataset with labeled examples (dropout or persistence) and learn to predict the outcome for new, unseen data. For this we will be using ***Logistic Regression***: A simple and interpretable model that calculates the probability of an event occurring. It's suitable for linearly separable data and it will also provide probabilities of dropout, making it easy to understand the likelihood of an event.

**References:**

**Most important primary resources :**

**1)**[**Surveys and statistical programs - Postsecondary Student Information System (PSIS) (statcan.gc.ca)**](https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=28999)

**2)** [**Over 1-in-4 (26%) Canadian post-secondary students said they have considered dropping out of their program because of money, new poll finds (newswire.ca)**](https://www.newswire.ca/news-releases/over-1-in-4-26-canadian-post-secondary-students-said-they-have-considered-dropping-out-of-their-program-because-of-money-new-poll-finds-831703839.html)

**3)**[**Regional-Data-Tables-For-Wesbite-Student-Poll-2024.pdf (embark.ca)**](https://www.embark.ca/wp-content/uploads/2024/08/Regional-Data-Tables-For-Wesbite-Student-Poll-2024.pdf)

**Secondary resources to be considered :**

* [Over 1-in-4 (26%) Canadian post-secondary students said they have considered dropping out of their program because of money, new poll finds (newswire.ca)](https://www.newswire.ca/news-releases/over-1-in-4-26-canadian-post-secondary-students-said-they-have-considered-dropping-out-of-their-program-because-of-money-new-poll-finds-831703839.html)
* [Student debt from all sources, by province of study and level of study (statcan.gc.ca)](https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3710003601)
* <https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention?resource=download>
* [https://open.alberta.ca/opendata?q=school HYPERLINK "https://open.alberta.ca/opendata?q=school&sort=score+desc&page=2"& HYPERLINK "https://open.alberta.ca/opendata?q=school&sort=score+desc&page=2"sort=score+desc HYPERLINK "https://open.alberta.ca/opendata?q=school&sort=score+desc&page=2"& HYPERLINK "https://open.alberta.ca/opendata?q=school&sort=score+desc&page=2"page=2](https://open.alberta.ca/opendata?q=school&sort=score+desc&page=2)
* <https://www150.statcan.gc.ca/n1/pub/81-582-x/81-582-x2024001-eng.htm>
* <https://www.weforum.org/events/sustainable-development-impact-meetings-2024/sessions/ai-for-global-good/>
* <https://www150.statcan.gc.ca/n1/pub/81-599-x/81-599-x2022002-eng.htm>
* <https://open.canada.ca/data/en/dataset/ffa27e47-bfce-4c55-98c0-5f3769316042>
* <https://www.weforum.org/agenda/2024/05/ways-ai-can-benefit-education/>
* <https://onlinedegrees.sandiego.edu/artificial-intelligence-education/>
* <https://phys.org/news/2024-06-ai-upper-secondary-dropout-early.html>
* <https://www.sciencedirect.com/getaccess/pii/S0957417423014355/purchase>
* <https://www.itu.int/hub/2022/11/predictive-ai-school-dropout-prevention-latin-america/>
* <https://www.uoc.edu/en/news/2023/209-AI-detects-students-at-risk-dropping-out>
* <https://www.nature.com/articles/s41598-024-63629-0>