Addressing Educational Disparities: Assessing the Gap for Indigenous Community

Shafaq Khan, Viutika Rathod, Abhirup Ranjan, Anika Anjum Una and Neel Manish Pandya

*School of Computer Science,University of Windsor, Ontario,Canada*

Keywords: Indigenous Knowledge, Indigenous Perspectives, Educational Disparities, Indigenous Students, NoSQL- MongoDB Database

Abstract: This research paper explores the integration of Indigenous knowledge and perspectives in education to address the educational disparities faced by Indigenous students in Canada [Change.org, 2023]. It proposes a manage- ment system utilizing a NoSQL-MongoDB database and logistic regression model to predict student dropout rates based on key factors such as Cultural Identity, Gender, Government Funding, and more [Government of Canada, 2021]. The logistic regression model achieved an accuracy rate of 0.831 on the testing dataset and provided valuable insights into the factors influencing dropout rates among Indigenous students. The paper emphasizes the importance of cultural sensitivity, ethical considerations, and collaboration with Indigenous communities throughout the research process. While logistic regression offers interpretability and simplicity, future work may explore the use of other machine learning models and qualitative data to enhance accuracy and gain deeper insights. The goal is to promote educational equity and inclusivity while respecting Indige- nous knowledge and aspirations in Canadian education.

# INTRODUCTION

The Indigenous communities in Canada experience significantly higher dropout rates compared to non- Indigenous communities in this part of the world, at- tributed to a range of factors such as historical and inter-generational impacts, financial inequities, so- cial marginalization, lack of support systems, and geographic barriers [Government of Ontario, 2017]. These factors restrict access to quality education for Indigenous students, resulting in lower educational achievements compared to the general population of Canada [CFS Ontario, 2021].

Outline of the problem of educational disparities faced by Indigenous students in Canada, there are multiple factors such as the legacy of colonialism, residential schools, racism, and insufficient funding hinder access to post-secondary education, which is recognized as a treaty right for Indigenous students. Funding inadequacy and access difficulties contribute to a significant disparity in educational opportuni- ties [Brown, 2023]. Also, the lack of necessary in- frastructure has hindered the distribution of cultur- ally relevant educational curricula to Indigenous com- munities. Teaching practices in non-Indigenous in- stitutions need refinement, focusing on incorporating Indigenous history, cultures, and perspectives, and addressing racism and marginalization [Government

of Canada, Interagency Advisory Panel on Research Ethics, 2019].

Our work in this research paper is motivated by a deep understanding of past injustices and current hardships faced by Indigenous communities. We aim to address the lack of cultural responsiveness in the mainstream educational system and promote justice, fairness, and reconciliation [Weston, 2019]. By em- bracing empathy and recognizing the generational ef- fects of policies like the residential school system, we seek to restore pride, dignity, and self-respect among Indigenous children [Kim, 2019]. Through cultural responsiveness, we aim to foster respect, understand- ing, and inclusive education for all students.

Our objective here is to bridge the educational gap between Indigenous and non-Indigenous populations in Canada by addressing issues such as insufficient funding, cultural disconnection, discrimination, and lack of support systems for Indigenous communities. We aim to assess and analyze the quality of educa- tion for Indigenous students, promote evidence-based decision making, and ensure educational equity. In- tegrating Indigenous knowledge systems, languages, and histories into the curriculum is a crucial focus, empowering Indigenous students and improving the educational experience for all.

To achieve our goals, we propose a model that uti- lizes a NoSQL database i.e. MongoDB, to address

educational disparities faced by Indigenous students. Also using the Logistic Model we will capture unique information such as cultural identity, language pro- ficiency, community involvement, gender, and other relevant details, while adhering to Indigenous data governance principles. This will assist us in deter- mining whether a particular person may drop out, and if so, it will enable concerned local or governmental bodies to take appropriate action. It will also enable comprehensive data integration from various sources, evaluating the success of educational initiatives and facilitating timely support and prevention of widening educational gaps. The chosen NoSQL database en- sures data privacy and security, aligning with ethical principles and relevant laws. There is a lack of pub- lished papers addressing our identified issue, making our research work unique. Through our comprehen- sive model and data-driven approach, we aim to con- tribute to the development of effective strategies and policies that empower Indigenous students and bridge the educational gap in Canada.

# BACKGROUND STUDY

This paper explores the role of educators, particu- larly those teaching aspiring conservation practition- ers, in responding to the Truth and Reconciliation Commission (TRC) and the National Inquiry into Missing and Murdered Indigenous Women and Girls (MMIWG). The emphasis is on the significance of reconciliation and ethical engagement with Indige- nous Peoples through a revolutionary approach to teaching indigenous knowledge. The focus is on fos- tering understanding, empathy, and respect for In- digenous knowledge and aspirations while equipping students with critical analysis skills for ethical en- gagement. The paper calls for universities to ”Indi- genize” their approaches, embracing anti-racism, hu- mility, reciprocity, and confronting ongoing colonial- ism and white supremacy. The goal is to create a learning environment that respects and values Indige- nous scholars, knowledge,and voices, fostering re- conciliatory relationships between conservation prac- titioners and Indigenous Peoples. Efforts should go beyond course content and focus on building an anti- racist, anti-oppressive campus culture that centers In- digenous perspectives and enables Indigenous intel- lectual expression. The paper advocates for hiring more Indigenous faculty members and centering In- digenous Peoples as experts about themselves in the curriculum. The goal is to prepare students to engage with Indigenous Peoples in a just and affirming man- ner while respecting Indigenous knowledge and aspi-

rations. The possible difficulties of successfully inte- grating Indigenous knowledge within the current cur- riculum and ensuring cultural sensitivity and authen- ticity in its application might, however, be a drawback of this strategy [Wu et al., 2019].

This paper focuses on the use of digital technolo- gies to support Indigenous people’s language and lit- eracy learning, particularly in English. Here, the em- phasis is on addressing the negative inter-generational impacts of colonization and socioeconomic stress on Indigenous academic performance. The system- atic review of 25 empirical studies provides insights into the efficacy of digital technology in support- ing Indigenous learners. While the studies demon- strate positive outcomes, there are limitations, such as the lack of rigorous research methods and com- prehensive reporting. To improve the effectiveness of digital technology-based interventions, future re- search should consider culturally relevant multilitera- cies frameworks, engage in longitudinal studies to track students’ progress and incorporate more Indige- nous cultural elements. Additionally, there is a need for data coding schemes to capture the nuances of In- digenous language and literacy learning. The research emphasizes the importance of culturally responsive practices, partnership with Indigenous communities, and addressing the unique contextual factors affecting Indigenous education. The reliance on digital infras- tructure and access to technology, however, may be a weakness of their strategy or approach and present difficulties in isolated or disadvantaged Indigenous communities where dependable internet connectivity may be scarce [Li et al., 2021].

This paper presents the NOW (Northern Oral Lan-

guage and Writing) Play project aiming to support young children’s oral language and writing skills in northern rural and Indigenous communities. The re- search emerged from the concerns raised by kinder- garten teachers in northern Canadian communities about students’ limited language abilities upon enter- ing school. The project involved collaborating with public school divisions in small rural communities and local education authorities in Indigenous regions of four Canadian provinces. Focusing on the assess- ment and support of young children’s oral language and writing development in play contexts. These communities face challenges related to geographic isolation, limited teaching resources, and a lack of op- portunities for teachers’ professional learning. The project emphasized the use of play contexts to en- hance children’s language and literacy skills. Addi- tionally, the project aimed to develop a culturally rel- evant oral language assessment tool. This tool was designed to capture the diverse ways in which chil- dren use language for various social purposes dur-

ing play and small-group academic activities. Teach- ers and researchers collaboratively analyzed video recordings and transcripts of children’s play to iden- tify different language uses and track individual chil- dren’s language development [Stagg Peterson and Dwyer, 2016]. However, the project also acknowl- edges challenges such as parental resistance to play based approaches and the necessity for cultural sen- sitivity in research methods. Careful navigation of these obstacles is crucial to ensure the project’s ul- timate success and positive impact.

All these studies focus on the integration of in- digenous knowledge across a range of educational contexts. These papers acknowledge the challenges and limitations in implementing the overall goal of integrating Indigenous knowledge and perspectives in education.

# PROPOSED MODEL

In this section, we propose a predictive model us- ing Logistic Regression, a binary classification algo- rithm. The model predicts whether a student will drop out (1) or not (0) by using the key factors that have been identified as input variables. Its interpretable re- sults allow for a clear understanding of the impact of predictor variables on dropout rates, helping to iden- tify key factors contributing to the disparities [Das, 2021]. Moreover, logistic regression provides prob- ability estimates, prioritizing higher-risk students for support systems. With its low complexity, minimal data preprocessing requirements, and ease of imple- mentation, logistic regression offers a practical and efficient solution for handling the data management system. While it may not capture complex relation- ships as effectively as advanced algorithms, combin- ing it with other techniques through ensembling could further enhance its predictive performance and lead to more comprehensive strategies for bridging the edu- cational gap among Indigenous communities.

The proposed model takes a systematic approach

to predicting the likelihood of dropout among Indige- nous students (as shown in Figure 1). It starts by compiling pertinent data from various sources, such as governmental agencies and academic institutions [Bisong, 2019].

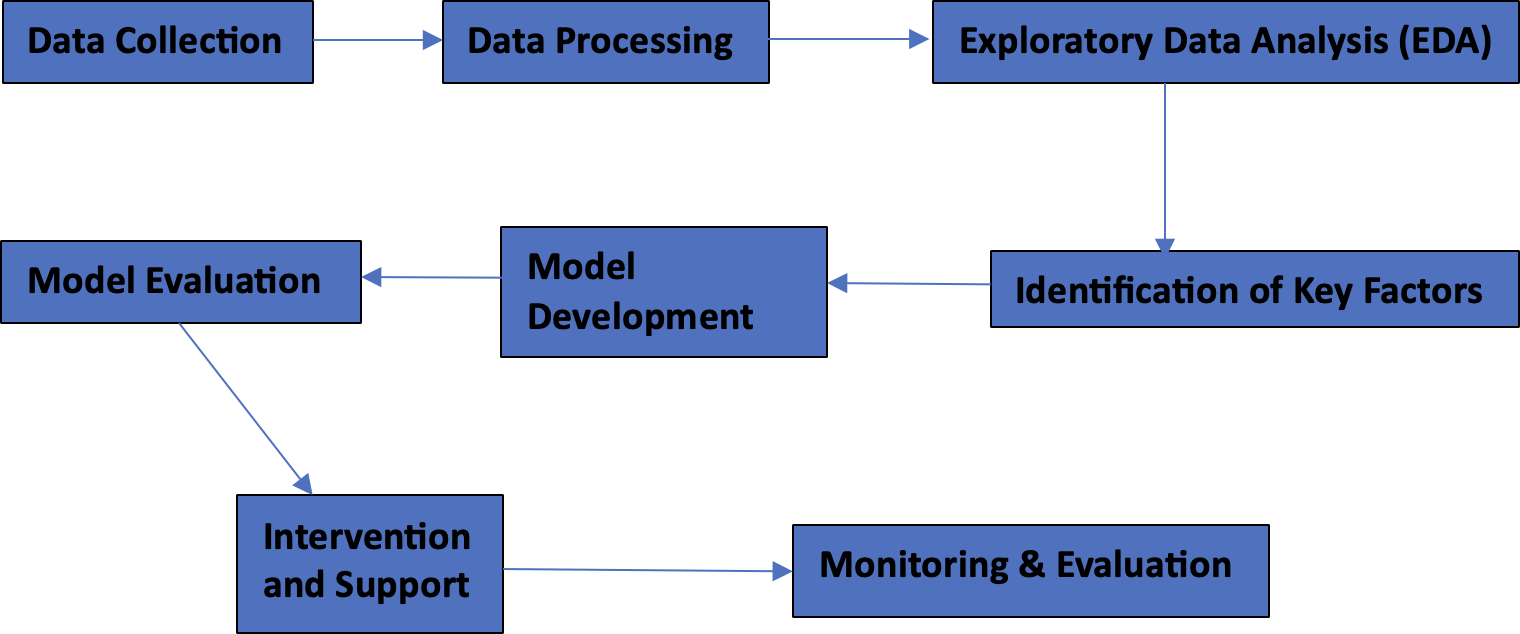


Figure 1: Workflow diagram.

Key traits like cultural identity, language profi- ciency, and educational level are chosen after the data has been preprocessed. Following that, the dataset is divided into training and test sets. To determine the correlation between the selected features and dropout outcomes, the logistic regression model is trained us- ing the training data. The model is tested on the test- ing set, and metrics like accuracy, precision, recall, F1-score, and ROC-AUC [K, 2020] are used to as- sess its performance. An effective predictive model for comprehending dropout patterns among Indige- nous students is being developed through this thor- ough process.

# METHODOLOGY AND EXPERIMENTATION

The logistic regression model was implemented as part of our proposed management system to address educational disparities faced by Indigenous students in Canada. The model’s objective was to predict the likelihood of student dropout (1) or non-dropout

(0) based on key factors, including Cultural Identity, Gender, Government Funding, Type of Educational Institute, Employment Sector, Language Proficiency, Community Involvement, Age, and Level of Educa- tion. We trained the model using a binary classifi- cation algorithm and evaluated its performance using various metrics.

## Data Collection

A variety of sources, including the Alberta Open Data Portal, Statistics Canada, the FNIGC (First Nations Information Governance Center), and the Govern- ment of British Columbia, were used to collect the data for data preprocessing in Indigenous education research. The dataset comprises diverse information, including demographic data, cultural identity, gen- der, government funding, enrollment figures, and aca- demic performance metrics. To ensure accuracy and suitability for analysis, the data underwent rigorous cleaning, transformation, and organization processes.

Outliers were addressed, features were normalized or encoded as required, and missing values were im- puted. Such preprocessing is crucial for subsequent analysis, such as logistic regression modeling, which aims to shed light on the variables influencing the aca- demic performance and dropout rates of Indigenous students.

The Figure 2 illustrates the credential they have achieved so far in the Alberta region based on the dataset available on Alberta Open Data Portal web- site.

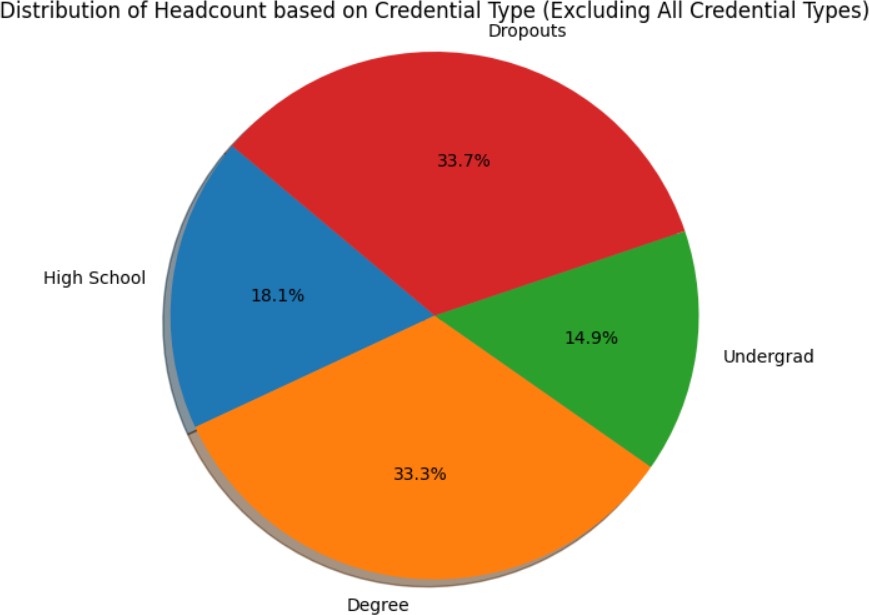


Figure 2: Distribution of Indigenous students based on their education level in Alberta Region.

## Data Processing

For data preprocessing, we will be categorizing the fields into two categories: Binary and Numerical variables [Rao et al., 2023]. After one-hot encoding, there will be only two values for categorical variables which we will include as binary variables. Here is the categorization:

### Binary Variables

* + - **Cultural Identity (after binary encoding)**
      * First Nations (1 for ”First Nations”, 0 for oth- ers)
      * Inuit (1 for ”Inuit”, 0 for others)
      * Me´tis (1 for ”Me´tis”, 0 for others)

### Gender (after binary encoding)

* + - * Gender (1 for ”Female”, 0 for ”Male”)

### Government Funding (after binary encoding)

* + - * Government Funding (1 for ”Yes”, 0 for ”No”)

### Numerical Variables

* + - **Language Proficiency (after label encoding)**
      * Language Proficiency (0 for ”Basic”, 1 for ”In- termediate”, 2 for ”Advanced”)

### Community Involvement (after label encoding)

* + - * Community Involvement (0 for ”Low”, 1 for ”Medium”, 2 for ”High”)

### Age (Numerical variable)

* + - * Age (numeric values)

### Level of Education (after label encoding)

* + - * Level of Education (0 for ”School”, 1 for ”High School”, 2 for ”Bachelor’s”, 3 for ”Master’s”, etc.)

### Type of Educational Institute (after label en- coding)

* + - * Type of Educational Institute (0 for ”Public School”, 1 for ”Private School”, 2 for ”Home- schooling”, 3 for ”Online Learning”)

### Employment Sector (after label encoding)

* + - * Employment Sector (0 for ”Government”, 1 for ”Private”, 2 for ”Others”)

Table 1: Summary of Categorical and Numerical Features in the Dataset.

|  |  |
| --- | --- |
| Field | Category |
| First Nations | Binary (Cultural Identity) |
| Inuit | Binary (Cultural Identity) |
| Me´tis | Binary (Cultural Identity) |
| Gender | Binary |
| Government Funding | Binary |
| Language Proficiency | Numerical (Ordinal) |
| Community Involvement | Numerical (Ordinal) |
| Age | Numerical (Ordinal) |
| Level Of Education | Numerical (Ordinal) |
| Type of Educational Insti-  tute | Numerical (Ordinal) |
| Employment Sector | Numerical (Ordinal) |

## Feature Selection

Recursive feature elimination (RFE) [Misra and Ya- dav, 2020] is used in the proposed algorithm for addressing educational disparities among Indigenous students in Canada to pinpoint the key characteristics that significantly influence the likelihood of dropout prediction made by logistic regression. Initially, the logistic regression model is trained using all perti- nent features, including binary and numerical vari- ables. Following a systematic elimination of less sig- nificant features, the RFE process ranks the remain- ing features according to how well they predict aca- demic outcomes. The subset of crucial elements with the greatest influence on Indigenous students’ dropout rates is identified through this iterative process.

Algorithm 1: Algorithm: Feature Selection using RFE and Logistic Regression.

**Data:** Input data *X* , Target variable *y* **Result:** Selected features *X*selected Encode categorical features;

*X*enc OneHotEncode(*X* );

*←*

Split data into training and testing sets;

*X*train*, X*test*, y*train*, y*test

*←*

split(*X*enc*, y,* test size = 0*.*2*,* random state = 42*,* stratify = *y*);

Create logistic regression model with

balanced class weights;

*model* LogisticRegression(class weight = ’balanced’);

*←*

Implement RFE for feature selection;

num features 5;

*←*

*r f e* RFE(*model,* n features to select = num features);

*←*

*r f e.*fit(*X*train*, y*train);

Fit model on selected features; *X*train sel *X*train[:*, r f e.*support ]; *model.*fit(*X*train sel*, y*train);

*←*

Predict target variable on the testing set;

*X*test sel *X*test[:*, r f e.*support ]; *y*pred *model.*predict(*X*test sel); Calculate F1-score and accuracy; *f* 1 F1-score(*y*test*, y*pred);

*←*

*←*

*←*

*accuracy* accuracy score(*y*test*, y*pred);

*←*

Print results;

**print**(”F1-score for ’Dropout’:”, f1);

**print**(”Accuracy:”, accuracy);

The F1-score is used in the above algorithm to evaluate how well the logistic regression model pre- dicted educational outcomes for Indigenous students. Particularly when working with datasets that are un- balanced, it aids in assessing the model’s capacity to balance precision and recall. The computational com- plexity of RFE, which is roughly O(*np*2*/*2), is also taken into account when choosing the features. With RFE’s iterative design, subsets of features are used to train the logistic regression model, allowing for the

its performance will be assessed using the test set

2. The split ratio is typically 80% training data and 20% test data, but it can be changed depending on the size of the dataset and the particular require- ments.

## Model Training

In order to accurately predict whether a student is likely to drop out (1) or not (0) based on the selected features: ”Cultural Identity,” ”Gender,” ”Government Funding,” ”Type of Educational Institute,” ”Employ- ment Sector,” ”Language Proficiency,” ”Community Involvement,” ”Age,” and ”Level of Education,” the logistic regression model must be trained to recog- nize patterns and relationships within the training data [Parker et al., 2013].

For each data point in the training set, the calcu- lated values in the context of logistic regression re- fer to the linear combinations of the chosen features (’Cultural Identity’, ’Gender’, ’Government Fund- ing’, ’Type of Educational Institute’, ’Employment Sector’, ’Language Proficiency’, ’Community In- volvement’, ’Age’, ’Level of Education’). For each data point, the model will compute a weighted sum of the feature values and add a bias term. The linear combination (y) for a single data point (x) is com- puted mathematically as follows:

*y* = *b*0 + *b*1*x*1 + *b*2*x*2 + *. . .* + *bnxn* (1)

### Here:

* + - y is the calculated value for a given data point.
    - b0 is the bias term (intercept).
    - b1, b2, ..., bn are the weights (coefficients) as- signed to each feature.
    - x1, x2, ..., xn are the feature values for the corre- sponding data point.

The calculated values (y) will then be subjected to the sigmoid function to convert them into probabili- ties. The formula for the function is:

effective identification of key predictors to reduce ed- ucational disparities among Indigenous communities.

## 4.4 Data Split

The ”train-test split,” in which the dataset is divided into two parts: “**a training set**” and “**a test set**”, is the most popular data splitting technique.

1. The training set and the test set should be sep- arated from the dataset. The logistic regression model will be trained using the training set, and

*p* = 1

1 + exp(*−y*)

(2)

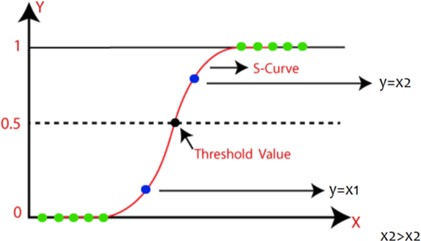


Figure 3: The S Curve: Visualization of Logistic Regression’s Probabilistic Model.

Thus, p is the logistic regression model’s out- put—the likelihood that a student will drop out—for a specific data point.

The exponential representation of -y is represented by *exp−y*. No matter the range of the calculated values (y), the sigmoid function (as shown in Figure **??** [Es- sampally, 2020]) will make sure that the output prob- abilities (p) are bounded between 0 and 1. For the corresponding data point, the model will predict a dropout (1) when p is greater than or equal to 0.5, and a non-dropout (0) when p is less than or equal to 0.5. The regression algorithm will iteratively adjust the weights (b1, b2,..., bn) and the bias term (b0) during the model training process using optimization tech- niques like gradient descent. The goal is to identify the best combination of weights and biases to mini- mize the discrepancy between the predicted probabil- ities and the actual binary outcomes (dropout or non- dropout) in the training data, thereby enhancing the model’s capacity to make precise predictions on fresh, untested data.

## Model Evaluation

Evaluation metrics like accuracy, precision, recall, F1-score, and ROC-AUC (Area Under the Curve - Re- ceiver Operating Characteristic) are used to assess the performance of the logistic regression model, which was trained using the features like Cultural Identity, Gender, Government Funding, Type of Educational Institute, Employment Sector, Language Proficiency, Community Involvement, Age, and Level of Educa- tion. These metrics evaluate how well the model per- forms in accurately predicting Indigenous students’ likelihood of dropping out of school. We can assess the model’s efficacy in addressing educational dispar- ities and promoting educational equity among Indige- nous communities in Canada by analyzing these eval- uation results on the testing dataset.

## Interpretability

The logistic regression model’s feature importance (coefficients) offers important insights into how each feature affects the likelihood that Indigenous students will drop out of school. While negative coefficients suggest factors linked to lower dropout rates, pos- itive coefficients point to factors linked to higher dropout rates. Stakeholders can pinpoint key causes of educational disparities by looking at these coef- ficients for categories like Cultural Identity, Gender, Government Funding, Type of Educational Institute, Employment Sector, Language Proficiency, Commu- nity Involvement, Age, and Level of Education. In order to promote educational equity for Indigenous communities, targeted interventions and support sys- tems are informed by this understanding. While there are other interpretability techniques, Feature Impor- tance (Coefficients) stands out for its clarity, usabil- ity, and universal interpretability, enabling evidence- based decision-making to effectively close the educa- tional gap for Indigenous students.

# RESULTS

The logistic regression model demonstrated promis- ing results in predicting dropout rates among Indige- nous students. The model achieved an accuracy rate of 0.831 on the testing dataset, indicating its ability to correctly classify dropout outcomes. Moreover, the F1 score was 0.831 (as shown in Figure 4), showcas- ing its effectiveness in identifying students at risk of dropping out and minimizing false positives.

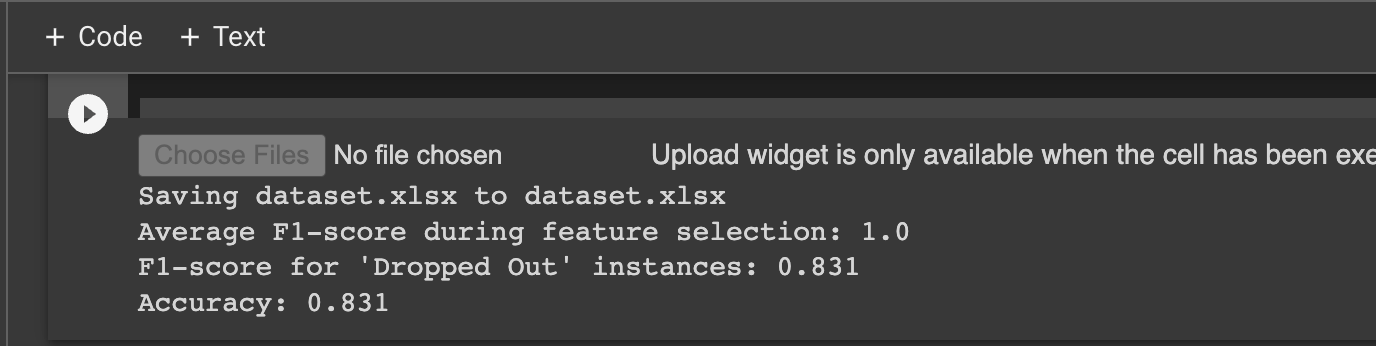


Figure 4: Showing the Accuracy for Dropped out instances for Indigenous Students.

The interpretability of the logistic regression model allowed us to identify key factors influencing dropout rates among Indigenous students. The coeffi- cients for various features provided valuable insights into the impact of Cultural Identity, Gender, Gov- ernment Funding, Type of Educational Institute, Em- ployment Sector, Language Proficiency, Community Involvement, Age, and Level of Education on edu- cational outcomes. For instance, positive coefficients indicated factors associated with higher dropout rates, while negative coefficients pointed to factors linked to lower dropout rates. This understanding empow-

ered us to design targeted interventions and support systems to enhance educational equity and bridge the educational gap for Indigenous students in Canada.

# LIMITATIONS/CHALLENGES

While the proposed approach to address educational disparities faced by Indigenous students in Canada through the utilization of a NoSQL database (Mon- goDB) and logistic regression model is promising, there are a few limitations that must be considered.

Initially, our approach was highly data-driven and relied on the quantity and availability of data. The in- digenous communities we are working on along with their historical factors, privacy concerns, and limited data collection impose challenges in obtaining correct and descriptive data. Missing values in features or target variables can lead to biased estimates and inac- curate predictions, even with imputation techniques, which may introduce biases of their own. Addition- ally, biased data with over/underrepresented groups can result in unfair and discriminatory predictions, re- quiring proactive measures to address data bias. Out- liers, as extreme values, can exert undue influence on model parameters and should be properly detected and handled for improved accuracy. Moreover, sam- pling bias arising from non-random or unrepresenta- tive data collection can hinder model generalization to new data. Data heterogeneity stemming from di- verse sources may necessitate harmonization efforts to ensure meaningful analysis.

Canada’s Indigenous communities encompass a

rich tapestry of diversity, comprising over 600 rec- ognized First Nations, Me´tis, and Inuit communities, each with distinct cultural and socio-economic back- grounds. Acknowledging and comprehending these complexities are essential in addressing the educa- tional disparities faced by Indigenous students. Cul- tural diversity plays a pivotal role as communities may prioritize traditional knowledge and language preservation, while others seek to integrate modern education with their heritage. Socioeconomic factors add another layer of variation, with economic realities varying between remote communities, where oppor- tunities may be limited, and urban populations fac- ing unique challenges that influence educational ac- cess and outcomes. Historical traumas, like coloniza- tion, continue to impact educational engagement and demonstrate resilience in preserving their cultures. Language and identity are intertwined and directly influence educational experiences. Moreover, access to educational resources differs across communities. Disparities also diverge between remote and urban In-

digenous communities, driven by geographic barriers and infrastructure challenges.

Addressing resource constraints requires strate- gies and collaborative efforts. Limited funding poses challenges for educational initiatives, data manage- ment systems, and advanced analytics, alongside the high costs of technology infrastructure and skilled personnel. Moreover, technical constraints like the lack of specialized expertise, unreliable internet con- nectivity in remote communities, and resistance to technological changes add complexity.

In addition, our model, logistic regression, may provide insights into factors contributing to dropout rates, but it might not capture all the complexities of the educational disparities faced by Indigenous stu- dents. The model assumes linear relationships, but educational disparities may involve non-linear, inter- active, or complex patterns that it may not capture ef- fectively. One major drawback is its limited ability to capture nonlinear decision boundaries, which may not adequately represent situations influenced by mul- tiple factors. Additionally, it might overlook impor- tant feature interactions, like cultural identity or com- munity involvement, leading to oversimplified predic- tions. Handling high-dimensional data can be chal- lenging due to issues like multicollinearity and over- fitting. Its reliance on numeric features also makes it difficult to handle non-numeric data, such as cate- gorical or textual variables, without increasing model complexity. Moreover, there exists a trade-off be- tween model interpretability and predictive accuracy, with logistic regression being interpretable but poten- tially sacrificing performance in complex educational scenarios.

In addressing educational disparities, incorporat-

ing other machine learning models such as SVM, decision trees, clustering, and statistical techniques like correlation analysis and multiple regression anal- ysis can offer a more comprehensive understanding of the complexities involved and enhance predictive ac- curacy based on dataset characteristics and research goals. In conducting research in Indigenous educa- tion, upholding ethical principles is paramount. This includes obtaining informed consent and safeguard- ing sensitive data to protect confidentiality, while also respecting Indigenous values and involving com- munities in decision-making processes. Researchers must be cautious not to perpetuate stereotypes or stig- matize communities and demonstrate cultural sensi- tivity by acknowledging traditions and historical con- text. Transparency and accountability are essential, in documenting methodology and sharing findings with communities. The focus should be on benefiting In- digenous communities, aligning research outcomes with their goals, and ensuring the long-term sus-

tainability of data systems. Empowering Indigenous voices and addressing power imbalances in research is vital for promoting inclusivity and driving positive change in Indigenous education [Oguamanam, 2019].

# CONCLUSION AND FUTURE WORK

The future work of this research paper holds promis- ing directions to enhance its impact and tackle ad- ditional challenges in Indigenous education. Ensur- ing data privacy and security while collaborating with Indigenous communities is essential for comprehen- sive data collection. Cultural sensitivity should be prioritized throughout the research process to respect community values and align research outcomes with their needs. Incorporating qualitative data alongside quantitative measures can provide deeper insights into Indigenous students’ experiences. Exploring various machine learning models and forming collaborations between academic institutions, government agencies, and Indigenous communities can lead to more accu- rate predictions and sustainable change. Ultimately, future research should focus on promoting equitable and inclusive education while addressing educational disparities faced by Indigenous students. In con- clusion, the logistic regression model proved to be a valuable tool in addressing educational disparities faced by Indigenous students. Its accuracy and in- terpretability allowed for the identification of signif- icant predictors of dropout, enabling the develop- ment of evidence-based strategies for educational im- provement. By incorporating Indigenous data gover- nance principles and utilizing MongoDB as a NoSQL database, our management system offers a unique and comprehensive solution to empower Indigenous com- munities and promote educational equity in Canada. The integration of Indigenous knowledge systems, languages, and histories into the curriculum further enhances the educational experience for all students and fosters reconciliation. With our research and data-driven approach, we aim to contribute to the de- velopment of effective policies and strategies that pro- mote justice, fairness, and inclusivity in education for Indigenous communities.

# REFERENCES

Bisong, E. (2019). Logistic regression. pages 243–250. Brown, D. (2023). Ontario has ‘come quite far’ on indige-

nous education but there’s much more to be done: re-

port. *CBC*.

CFS Ontario (2021). Post secondary education and treaty rights.

Change.org (2023). Sign the petition.

Das, A. (2021). Logistic regression. pages 1–2. Essampally, D. (2020). Logistic regression. *Medium*.

Government of Canada (2021). Truth and reconciliation commission of canada.

Government of Canada, Interagency Advisory Panel on Re- search Ethics (2019). Tri-council policy statement: Ethical conduct for research involving humans - tcps 2 (2018) - chapter 9: Research involving the first na- tions, inuit and me´tis peoples of canada.

Government of Ontario (2017). Aboriginal students - col- leges and universities.

K, G. M. (2020). Machine learning basics: Logistic regres- sion.

Kim, P. J. (2019). Social determinants of health inequities in indigenous canadians through a life course ap- proach to colonialism and the residential school sys- tem. 3(1):378–381.

Li, J., Brar, A., and Roihan, N. (2021). The use of digi- tal technology to enhance language and literacy skills for indigenous people: A systematic literature review. *Computers and Education Open*, 2.

Misra and Yadav (2020). Improving the classification ac- curacy using recursive feature elimination with cross- validation. 11(3).

Oguamanam, C. (2019). Indigenous data sovereignty: Re- tooling indigenous resurgence for development.

Parker, P. D., Bodkin-Andrews, G., Marsh, H. W., Jerrim, J., and Schoon, I. (2013). Will closing the achieve- ment gap solve the problem? an analysis of primary and secondary effects for indigenous university entry. *Journal of Sociology*.

Rao, K. M., Saikrishna, G., and Supriya, K. (2023). Data preprocessing techniques: emergence and selection towards machine learning models - a practical review using hpa dataset. *Multimedia Tools and Applications*.

Stagg Peterson, S. and Dwyer, B. (2016). Research in canada’s northern rural and indigenous communities: Supporting young children’s oral language and writ- ing. *The Reading Teacher*, 70.

Weston, D. (2019). Importance of safe water for aboriginal children’s education. *Heart and Art*.

Wu, J., Lin, S., Kong, H., and Shi, H. (2019). The com- bination forecasting model of telecommunication user tricking account overdraft limit based on logistic re- gression and svm. pages 411–415.