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, 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 management system utilizing a 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 Indigenous knowledge and aspirations in Canadian education. This will help the local and federal authorities in determining and early forecasting that the student might drop out and accordingly the authorities can take action. Also the data and results from different regions can be help in determining how they are performing and if their action planning can be implemented in the other region as well to reduce the drop out ratio. This will help in the overall development of the indigenous community.

1 INTRODUCTION

The Indigenous communities in Canada experience significantly higher dropout rates compared to non-Indigenous communities in this part of the world, attributed to a range of factors such as historical and inter-generational impacts, financial inequities, social 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 opportunities [Brown, 2023]. Also, the lack of necessary infrastructure has hindered the distribution of culturally relevant educational curricula to Indigenous communities. Teaching practices in non-Indigenous institutions 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].Our goal is to assist the government authorities of Canada in understanding the dropout rates of Indigenous students so that they can devise appropriate strategies for their advancement and development. By embracing empathy and recognizing the generational effects 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, understanding, 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 education for Indigenous students, promote evidence-based decision making, and ensure educational equity. Integrating 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 utilizes a database to address educational disparities faced by Indigenous students. Also using the Logistic Model we will capture unique information such as cultural identity, language proficiency, community involvement, gender, and other relevant details, while adhering to Indigenous data governance principles. This will assist us in determining 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. There is a lack of published papers addressing our identified issue, making our research work unique. Through our comprehensive model and data-driven approach, we aim to contribute to the development of effective strategies and policies that empower Indigenous students and bridge the educational gap in Canada.

2 BACKGROUND STUDY

The paper by Wu, J., Lin, S., Kong, H., and Shi, H. (2019) titled "The Combination Forecasting Model of Telecommunication User Tricking Account Overdraft Limit Based on Logistic Regression and SVM explores the role of educators, particularly those teaching aspiring conservation practitioners, 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 Indigenous Peoples through a revolutionary approach to teaching indigenous knowledge. The focus is on fostering understanding, empathy, and respect for Indigenous knowledge and aspirations while equipping students with critical analysis skills for ethical engagement. The paper calls for universities to "Indigenize" their approaches, embracing anti-racism, humility, reciprocity, and confronting ongoing colonialism and white supremacy. The goal is to create a learning environment that respects and values Indigenous scholars, knowledge,and voices, fostering re-conciliatory relationships between conservation practitioners and Indigenous Peoples. Efforts should go beyond course content and focus on building an anti-racist, anti-oppressive campus culture that centers Indigenous perspectives and enables Indigenous intellectual expression. The paper advocates for hiring more Indigenous faculty members and centering Indigenous Peoples as experts about themselves in the curriculum. The goal is to prepare students to engage with Indigenous Peoples in a just and affirming manner while respecting Indigenous knowledge and aspirations. The possible difficulties of successfully integrating Indigenous knowledge within the current curriculum and ensuring cultural sensitivity and authenticity in its application might, however, be a drawback of this strategy [Wu et al., 2019].

In this study by Li, J., Brar, A., and Roihan, N. (2021) "The Use of Digital Technology to Enhance Language and Literacy Skills for Indigenous People: A Systematic Literature Review,"the focus is on the use of digital technologies to support Indigenous people's language and literacy learning, particularly in English. Here, the emphasis is on addressing the negative inter-generational impacts of colonization and socioeconomic stress on Indigenous academic performance. The systematic review of 25 empirical studies provides insights into the efficacy of digital technology in supporting Indigenous learners. While the studies demonstrate positive outcomes, there are limitations, such as the lack of rigorous research methods and comprehensive reporting. To improve the effectiveness of digital technology-based interventions, future research should consider culturally relevant multiliteracies frameworks, engage in longitudinal studies to track students' progress and incorporate more Indigenous cultural elements. Additionally, there is a need for data coding schemes to capture the nuances of Indigenous 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 infrastructure 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].

Exploring the NOW (Northern Oral Language and Writing) Play project, Stagg Peterson, S., and Dwyer, B. (2016) in their paper titled "Research in Canada's Northern Rural and Indigenous Communities: Supporting Young Children's Oral Language and Writing" shed light on efforts to enhance oral language and writing skills among young children in northern rural and Indigenous communities. The research

emerged from the concerns raised by kindergarten teachers in northern Canadian communities about students' limited language abilities upon entering 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 assessment 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 opportunities for teachers' professional learning. The project emphasized the use of play contexts to enhance children's language and literacy skills. Additionally, the project aimed to develop a culturally relevant oral language assessment tool. This tool was designed to capture the diverse ways in which children use language for various social purposes during play and small-group academic activities. Teachers and researchers collaboratively analyzed video recordings and transcripts of children's play to identify different language uses and track individual children's language development [Stagg Peterson and Dwyer, 2016]. However, the project also acknowledges challenges such as parental resistance to play based approaches and the necessity for cultural sensitivity in research methods. Careful navigation of these obstacles is crucial to ensure the project's ultimate success and positive impact.

The paper, "Educational Equity in Canada: The Case of Ontario's Strategies and Actions to Advance Excellence and Equity for Students" [Campbell, 2020] by Carol Campbell, delves into the educational equity initiatives in Ontario, Canada, focusing on two main strands: the Literacy and Numeracy Strategy and the Student Success/Learning to 18 Strategy. With a commitment to multiculturalism and diverse outcomes, the paper discusses policies targeting gender, language learners, special education needs, and Indigenous students within the context of two specific policy strands. While the literature review provides a comprehensive historical perspective, it could benefit from a more critical analysis of the strategies' effectiveness, including a nuanced exploration of Indigenous issues and a discussion of potential limitations. The paper considers students across K-12 levels but could be strengthened by specifying the age groups within this range.

Moules et al.'s (2014) [Schaub, 2020] research delves into the educational disparity between Indigenous and non-Indigenous students in Canada, employing hermeneutics for an in-depth understanding. Referring to it as an "educating gap," the study emphasizes the potential for actionable solutions. Examining historical issues in Indigenous education, the paper advocates for meaningful change while recognizing leadership challenges. The investigation explores the gap from various perspectives, including quantitative data, personal narratives, and perceptual insights. Despite the strengths in its approach, the study could strengthen its discussion by explicitly addressing limitations and providing a detailed analysis of the term "deficit model" in education. The authors acknowledge constraints such as time, financial resources, personal boundaries, and regional focus in coastal British Columbia, expressing confidence in the broader applicability of their findings.

The studies by Wu et al. (2019), Li et al. (2021), Stagg Peterson and Dwyer (2016),Campbell (2020) and Moules et al's(2014) collectively focus on the integration of Indigenous knowledge across various educational contexts. While acknowledging challenges and limitations in implementing this integration goal, these papers emphasize the importance of reconciliatory relationships, culturally responsive practices, and addressing unique contextual factors affecting Indigenous education. Notably, it should be highlighted that the background studies, specifically concentrate on younger students, offering insights into enhancing language and literacy skills among children in northern rural and Indigenous communities.

3 PROPOSED MODEL

In this section, we propose a predictive model using Logistic Regression, a binary classification algorithm. 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 results allow for a clear understanding of the impact of predictor variables on dropout rates, helping to identify key factors contributing to the disparities [Das, 2021]. Moreover, logistic regression provides probability estimates, prioritizing higher-risk students for support systems. With its low complexity, minimal data preprocessing requirements, and ease of implementation, logistic regression offers a practical and efficient solution for handling the data management system. While it may not capture complex relationships as effectively as advanced algorithms, combining it with other techniques through ensembling could further enhance its predictive performance and lead to more comprehensive strategies for bridging the educational gap among Indigenous communities.

The proposed model takes a systematic approach to predicting the likelihood of dropout among Indigenous students (as shown in Figure 1). It starts by compiling pertinent data from various sources, such as governmental agencies and academic institutions

[Bisong, 2019].

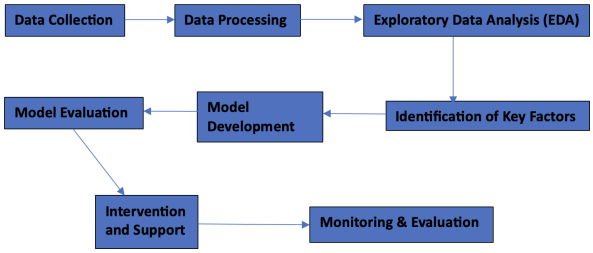

Figure 1: Workflow diagram.

Key traits like cultural identity, language proficiency, 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 using the training data. The model is tested on the testing set, and metrics like accuracy, precision, recall, F1-score, and ROC-AUC [K, 2020] are used to assess its performance. An effective predictive model for comprehending dropout patterns among Indigenous students is being developed through this thorough process.

4 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 Education. We trained the model using a binary classification algorithm and evaluated its performance using various metrics.

4.1 Data Collection

A variety of sources, including the Alberta Open Data Portal, Statistics Canada, the FNIGC (First Nations Information Governance Center), and the Government of British Columbia, contributed to the collection of data for preprocessing in Indigenous education research. The dataset, consisting of indigenous headcount enrolment within the Alberta Post Secondary Education-system with a sample size of 10000 records, includes various types of information, such as demographic data (e.g., age, gender), cultural identity, government funding status, enrollment figures, and academic 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 imputed. Such pre-processing steps are crucial, especially with a dataset of this size and complexity, to prepare the data for subsequent analysis, such as logistic regression modeling. This modeling aims to shed light on the variables influencing the academic performance and dropout rates of Indigenous students, providing valuable insights into the educational landscape.

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 website.

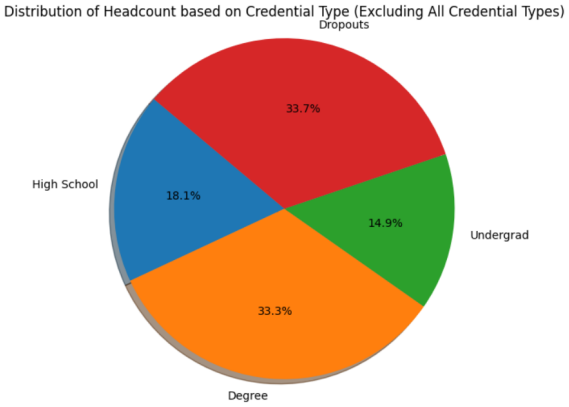

Figure 2: Distribution of Indigenous students based on their education level in Alberta Region.

4.2 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 others)
  - Inuit (1 for "Inuit", 0 for others)
  - Métis (1 for "Métis", 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 "Intermediate", 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 encoding)**
  - Type of Educational Institute (0 for "Public School", 1 for "Private School", 2 for "Homeschooling", 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)
Métis	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 Institute	Numerical (Ordinal)
Employment Sector	Numerical (Ordinal)

4.3 Feature Selection

Recursive feature elimination (RFE) [Misra and Yadav, 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 pertinent features, including binary and numerical variables. Following a systematic elimination of less significant features, the RFE process ranks the remaining features according to how well they predict academic 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;
$X_{\text{enc}} \leftarrow \text{OneHotEncode}(X)$;
Split data into training and testing sets;
$X_{\text{train}}, X_{\text{test}}, y_{\text{train}}, y_{\text{test}} \leftarrow$
$\text{split}(X_{\text{enc}}, y, \text{test\_size} = 0.2, \text{random\_state} = 42, \text{stratify} = y)$;
Create logistic regression model with balanced class weights;
$model \leftarrow \text{LogisticRegression}(\text{class\_weight} = \text{'balanced'})$;
Implement RFE for feature selection;
$\text{num\_features} \leftarrow 5$;
$rfe \leftarrow \text{RFE}(model, \text{n\_features\_to\_select} = \text{num\_features})$;
$rfe.\text{fit}(X_{\text{train}}, y_{\text{train}})$;
Fit model on selected features;
$X_{\text{train\_sel}} \leftarrow X_{\text{train}}[:, rfe.\text{support\_}]$;
$model.\text{fit}(X_{\text{train\_sel}}, y_{\text{train}})$;
Predict target variable on the testing set;
$X_{\text{test\_sel}} \leftarrow X_{\text{test}}[:, rfe.\text{support\_}]$;
$y_{\text{pred}} \leftarrow model.\text{predict}(X_{\text{test\_sel}})$;
Calculate F1-score and accuracy;
$f1 \leftarrow \text{F1-score}(y_{\text{test}}, y_{\text{pred}})$;
$accuracy \leftarrow \text{accuracy\_score}(y_{\text{test}}, y_{\text{pred}})$;
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 predicted educational outcomes for Indigenous students. Particularly when working with datasets that are unbalanced, it aids in assessing the model's capacity to balance precision and recall. The computational complexity 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 effective identification of key predictors to reduce educational 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 separated from the dataset. The logistic regression model will be trained using the training set, and its performance will be assessed using the test set. The training set is utilized to train the machine learning model. During this phase, the model learns patterns and relationships within the data. The test set is reserved for evaluating the model's performance. It consists of data that the model has not seen during the training phase, allowing for an unbiased assessment of its generalization capability.
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 requirements.Given our dataset size of 10000, we have allocated 8000 records for training and 2000 records for testing.

4.5 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,” ”Employment Sector,” ”Language Proficiency,” ”Community Involvement,” ”Age,” and ”Level of Education,” the logistic regression model must be trained to recognize patterns and relationships within the training data [Parker et al., 2013].

For each data point in the training set, the calculated values in the context of logistic regression refer to the linear combinations of the chosen features (’Cultural Identity’, ’Gender’, ’Government Funding’, ’Type of Educational Institute’, ’Employment Sector’, ’Language Proficiency’, ’Community Involvement’, ’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 computed mathematically as follows:

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n \quad (1)$$

Here:

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

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

$$p = \frac{1}{1+\exp(-y)} \quad (2)$$

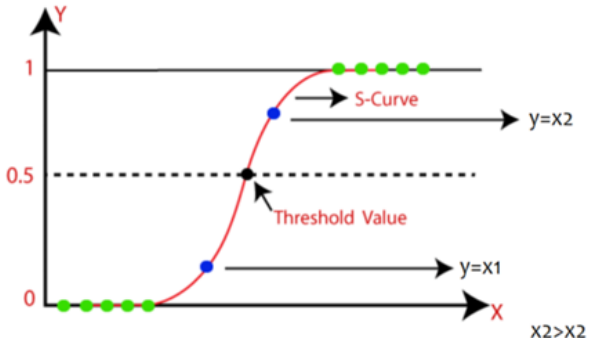

Figure 3: The S Curve: Visualization of Logistic Regression's Probabilistic Model.

Thus, p is the logistic regression model's output—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 **??** [Essampally, 2020]) will make sure that the output probabilities (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 techniques like gradient descent. The goal is to identify the best combination of weights and biases to minimize the discrepancy between the predicted probabilities 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.

4.6 Model Evaluation

Evaluation metrics like accuracy, precision, recall, F1-score, and ROC-AUC (Area Under the Curve - Receiver 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 Education. These metrics evaluate how well the model performs in accurately predicting Indigenous students' likelihood of dropping out of school. We can assess the model's efficacy in addressing educational disparities and promoting educational equity among Indigenous communities in Canada by analyzing these evaluation results on the testing dataset.

4.7 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, positive coefficients point to factors linked to higher dropout rates. Stakeholders can pinpoint key causes of educational disparities by looking at these coefficients for categories like Cultural Identity, Gender, Government Funding, Type of Educational Institute, Employment Sector, Language Proficiency, Community Involvement, Age, and Level of Education. In order to promote educational equity for Indigenous communities, targeted interventions and support systems are informed by this understanding. While there are other interpretability techniques, Feature Importance (Coefficients) stands out for its clarity, usability, and universal interpretability, enabling evidence-based decision-making to effectively close the educational gap for Indigenous students.

5 RESULTS

The logistic regression model demonstrated promising results in predicting dropout rates among Indigenous 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), showcasing its effectiveness in identifying students at risk of dropping out and minimizing false positives.

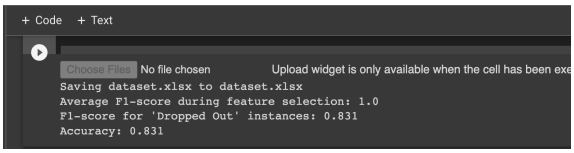

Figure 4: Showing the Accuracy for Dropped out instances for Indigenous Students.

The interpretability of the logistic regression model allowed us to identify key factors influencing dropout rates among Indigenous students. The coefficients for various features provided valuable insights into the impact of Cultural Identity, Gender, Government Funding, Type of Educational Institute, Employment Sector, Language Proficiency, Community Involvement, Age, and Level of Education on educational 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 empowered us to design targeted interventions and support systems to enhance educational equity and bridge the educational gap for Indigenous students in Canada.

6 LIMITATIONS/CHALLENGES

The proposed approach to addressing educational disparities faced by Indigenous students in Canada, utilizing a database and logistic regression model, shows promise but has several limitations. Initially, the approach was highly data-driven, relying on the quantity and availability of data, posing challenges due to privacy concerns, historical factors, and limited data collection in Indigenous communities. Missing values and biased data, along with outliers, can impact the accuracy of predictions, while sampling bias and data heterogeneity may hinder model generalization. Additionally, the logistic regression model, though insightful, may not capture all complexities, assuming linear relationships and struggling with high-dimensional and non-numeric data.

Understanding the rich tapestry of diversity among Canada's over 600 Indigenous communities is essential for addressing educational disparities effectively. Each community, comprising First Nations, Métis, and Inuit, possesses distinct cultural and socio-economic backgrounds. Factors such as cultural diversity, socioeconomic variations, historical traumas like colonization, and the intertwining of language and identity directly influence educational experiences. Disparities between remote and urban communities, driven by geographic barriers and infrastructure challenges, further complicate the educational landscape. Limited funding, technical constraints, and resistance to technological changes add to the challenges, requiring collaborative efforts and strategic initiatives to overcome resource constraints.

While the logistic regression model provides insights into factors contributing to dropout rates, it may not capture the full spectrum of complexities in Indigenous education. Its limitations in handling non-linear relationships, high-dimensional data, and non-numeric variables highlight the need for a more comprehensive approach. Incorporating various machine

learning models such as SVM, decision trees, and statistical techniques like correlation analysis can offer a broader understanding of educational disparities. Upholding ethical principles in Indigenous education research, including informed consent, confidentiality, cultural sensitivity, and community involvement, is crucial for ensuring positive, inclusive outcomes and promoting long-term sustainability.

7 CONCLUSION AND FUTURE WORK

The future work of this research paper holds promising directions to enhance its impact and tackle additional challenges in Indigenous education. Ensuring data privacy and security while collaborating with Indigenous communities is essential for comprehensive 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 accurate predictions and sustainable change. Ultimately, future research should focus on promoting equitable and inclusive education while addressing educational disparities faced by Indigenous students. In conclusion, the logistic regression model proved to be a valuable tool in addressing educational disparities faced by Indigenous students. Its accuracy and interpretability allowed for the identification of significant predictors of dropout, enabling the development of evidence-based strategies for educational improvement. By incorporating Indigenous data governance principles and utilizing MongoDB as a NoSQL database, our management system offers a unique and comprehensive solution to empower Indigenous communities 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 development of effective policies and strategies that promote 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 indigenous education but there's much more to be done: report. *CBC*.

Campbell, C. (2020). Educational equity in canada: the case of ontario's strategies and actions to advance excellence and equity for students. *School Leadership & Management*, 41(4-5):1–20.

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

Das, A. (2021). Logistic regression. pages 1–2.

Essampally, D. (2020). Logistic regression. *Medium*.

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

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

K, G. M. (2020). Machine learning basics: Logistic regression.

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

Li, J., Brar, A., and Roihan, N. (2021). The use of digital 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 accuracy using recursive feature elimination with cross-validation. 11(3).

Parker, P. D., Bodkin-Andrews, G., Marsh, H. W., Jerrim, J., and Schoon, I. (2013). Will closing the achievement 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*.

Schaub, A. (2020). Examining an apparent educating gap between non-indigenous and indigenous learners: A hermeneutic phenomenology approach.

Stagg Peterson, S. and Dwyer, B. (2016). Research in canada's northern rural and indigenous communities: Supporting young children's oral language and writing. *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 combination forecasting model of telecommunication user tricking account overdraft limit based on logistic regression and svm. pages 411–415.