Understanding Homelessness among Migrants to Thunder Bay using Machine Learning

Chandreen Ravihari[1], Vijay Mago[2], Rebecca Schiff[3] , and Ravi Gokani[4]

cliyanag@lakeheadu.ca, vmago@lakeheadu.ca, rschiff@lakeheadu.ca, rgokani@lakeheadu.ca

Lakehead University, Canada

Event: 5h Annual Canadian Homelessness Data Sharing Conference,
Data That Makes A Difference, University of Calgary.
May 18th-19th , 2022

Lakehead
UNIVERSITY

SSHRC CRSH

DATALab

Acknowledgement

- This is a SSHRC PEG-funded (2021-2022) collaborative project between Lakehead University (LU) and the District of Thunder Bay Social Services Administration Board (TBDSSAB), the largest social services agency in the District of Thunder Bay. SSHRC project # 892-2021-0014

Structure of the Presentation

- Research Questions

- The Dataset

- Methodology

  - Data Collection

  - Stages of Development

  - Data Cleaning and Preparation

  - Feature Selection

  - Synthetic Data

  - Machine Learning Models

- Results

- Discussion

Research Questions
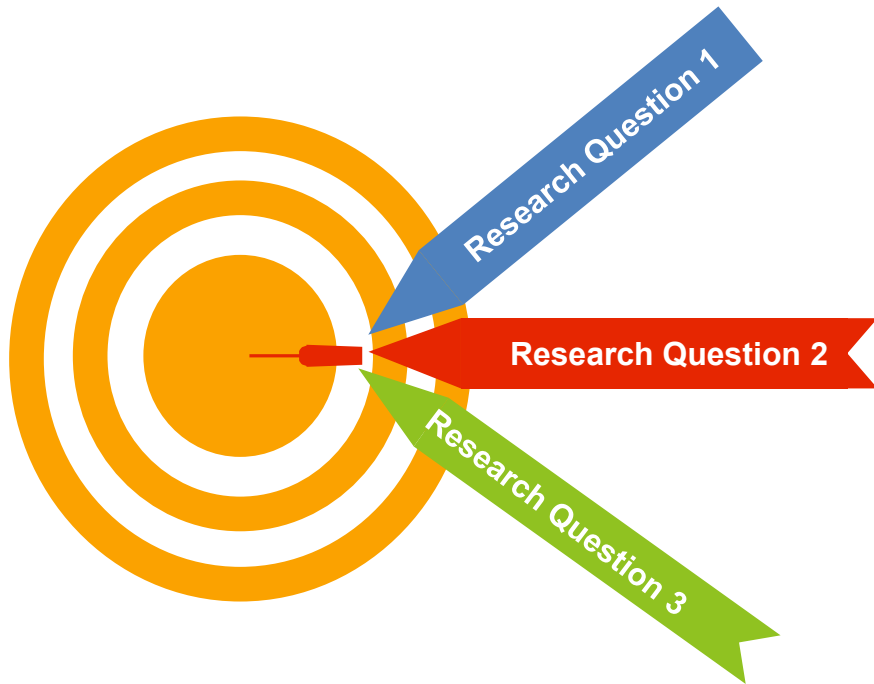
- Research Question 1:
  - What **factors** predict if an individual **stays or leaves** Thunder Bay?

- Research Question 2:
  - If an individual stays, **how long** are they likely to stay?

- Research Question 3:
  - What **factors** predict **stay duration**?

Research Question 1

Research Question 2

Research Question 3

The Dataset: Features used for the Study

- The dataset was collected from homeless adults through a survey conducted every Friday or Saturday at homeless shelters from 29th October 2021 to 9th April 2022
  - Shelter House, Salvation Army, Urban Abbey
  - Data collected by Graduate Student Research Assistants from the School of Social Work and Department of Health Sciences
- The shelter stay details (booked in, booked out dates) were gathered from TBDSSAB (HIFIS) databases
- Total number of records : 151

SSHRC ━ CRSH

Methodology: Data Collection over Time

Total Records over Time
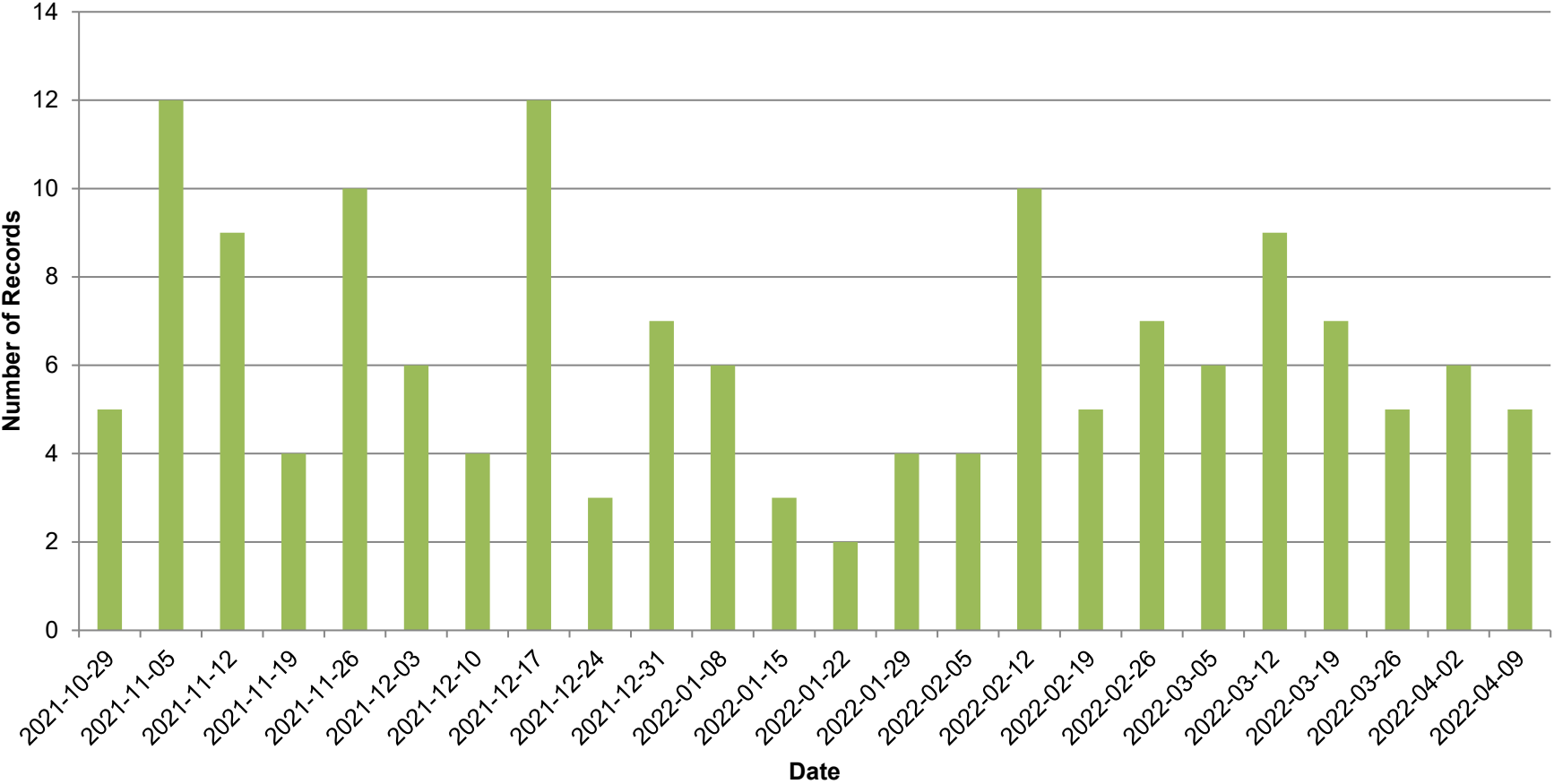
Methodology: Stages of Development
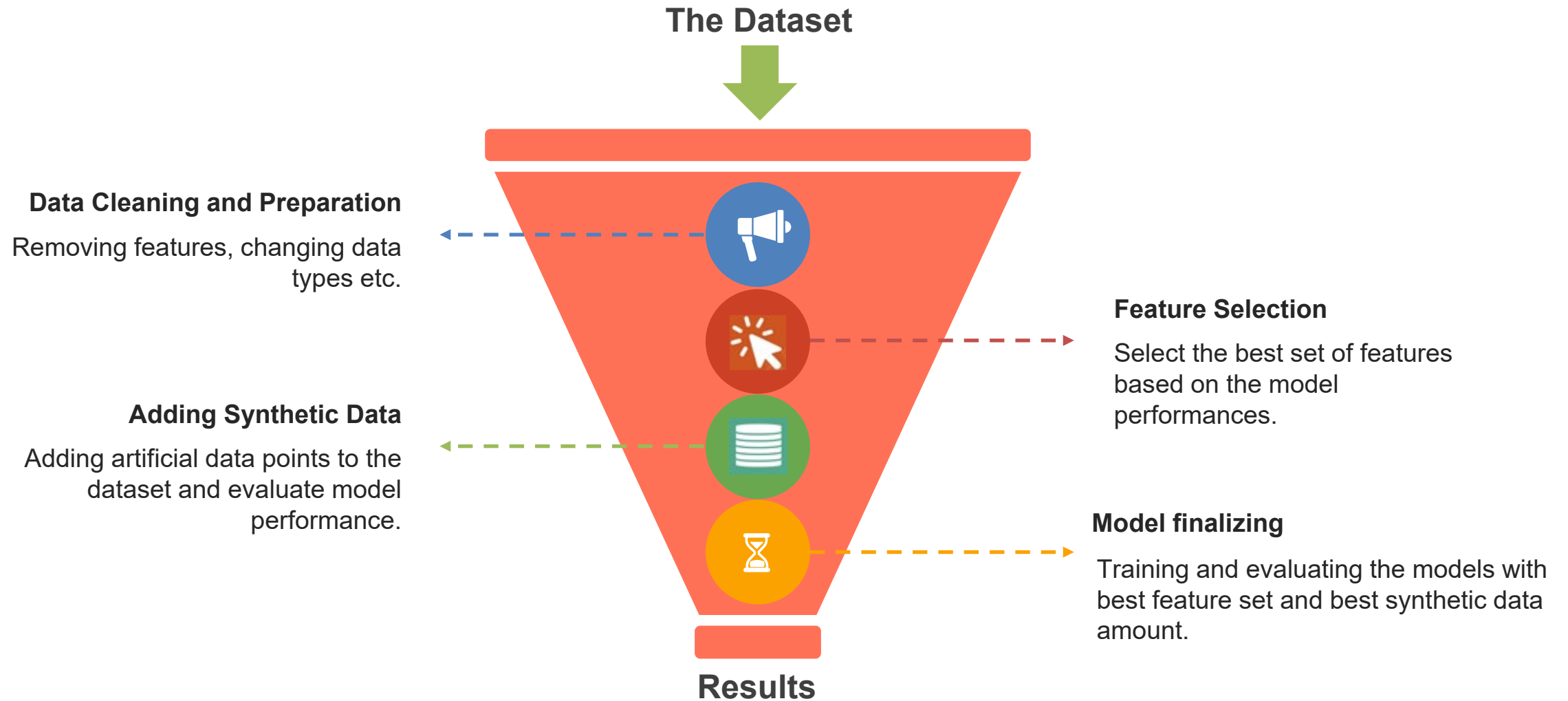
Methodology: Data Cleaning and Preparation

Total number of features in the dataset	33
Number of features used for the study	28
Removed features	First Name - Irrelevant
	Last Name - Irrelevant
	At what age did you come for schooling?– High NULL rate
	What type of schooling brought you here? – High NULL rate
	(Do you feel you can't go back to your community for any other reasons?) If so, what are the reasons? – Short Text
Generated features	Current Stay (Yes/No)
	Stay Duration
Numeric to categorical conversion	Age: 3 categories Stay Duration : 3 categories

Lakehead UNIVERSITY

SSHRC CRSH

DATALab

Methodology: Data Cleaning and Preparation

Predictive Class Label Generation

- Predicting variable: **Current stay (yes/no)**

- if the duration not at shelters >= 8 weeks;

    stay=no

  else

    stay =yes

- Predicting variable: **Stay duration**

- Predictive classes : 03 (1-20,   21-58,   59-245 days)

- Based on frequency distribution

Methodology: Best Feature Selection

Factor Selection was conducted using two approaches [1]

1. A filter-based approach
   - Using a **statistical test**, measure the relevance of features by their **correlation with dependent variable** (current stay/ stay duration)

1. A Wrapper-based approach
   - Use a **machine learning algorithm to select the most relevant subset of features** based on the performance of a ML model training.

[1] https://www.datasciencesmachinelearning.com/2019/10/feature-selection-filter-method-wrapper.html

SSHRC CRSH

Lakehead
UNIVERSITY

DATALab

Methodology: Adding Synthetic Data

- Performance of each model was measured by adding synthetic data from 1% to 50% of the total dataset.

SSHRC CRSH

Methodology: Types of Models Developed

1. Decision Tree

2. Support Vector Machine (SVM)

3. K-Nearest Neighbor (KNN)

4. Naive Bayes

5. Random Forest

6. Logistic Regression

7. Multi-Layer Perceptron (MLP)

8. Gradient Boosting Tree

Lakehead UNIVERSITY

SSHRC CRSH

DATALab

Results: RQ1 - Factors predict if an individual stays or leaves Thunder Bay?

26th February
Data Points: 119

Best Predictors

- Have you been hospitalized in Thunder Bay?
- Have you received any support for drug and/or alcohol use recently?
- Do you have family or friends in Thunder Bay?
- Did you find employment when you arrived in Thunder Bay?

26th March
Data Points: 140

Best Predictors

- Have you been hospitalized in Thunder Bay?
- Do you have family or friends in Thunder Bay?
- Was your family or were your friends a reason you came to Thunder Bay?
- Was mental health support a reason you came to Thunder Bay?
- What is the highest level of education you have completed?
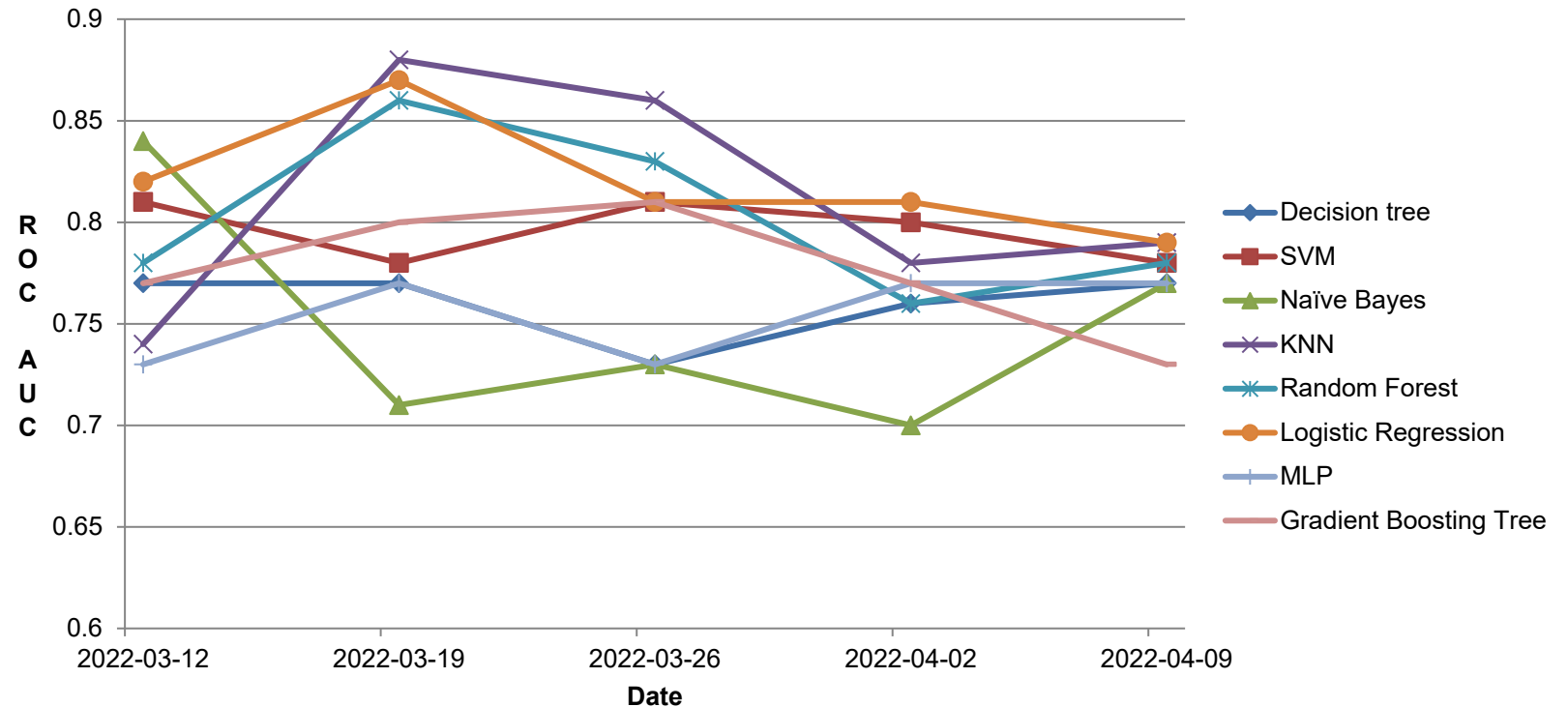
9th April
Data Points: 151

Best Predictors

- Do you have family or friends in Thunder Bay?
- Was education/school a reason you came to Thunder Bay?
- Was your family or were your friends a reason you came to Thunder Bay?
- What is the highest level of education you have completed?
- Have you been hospitalized in Thunder Bay?

SSHRC CRSH

LAKEHEAD UNIVERSITY

DATALab

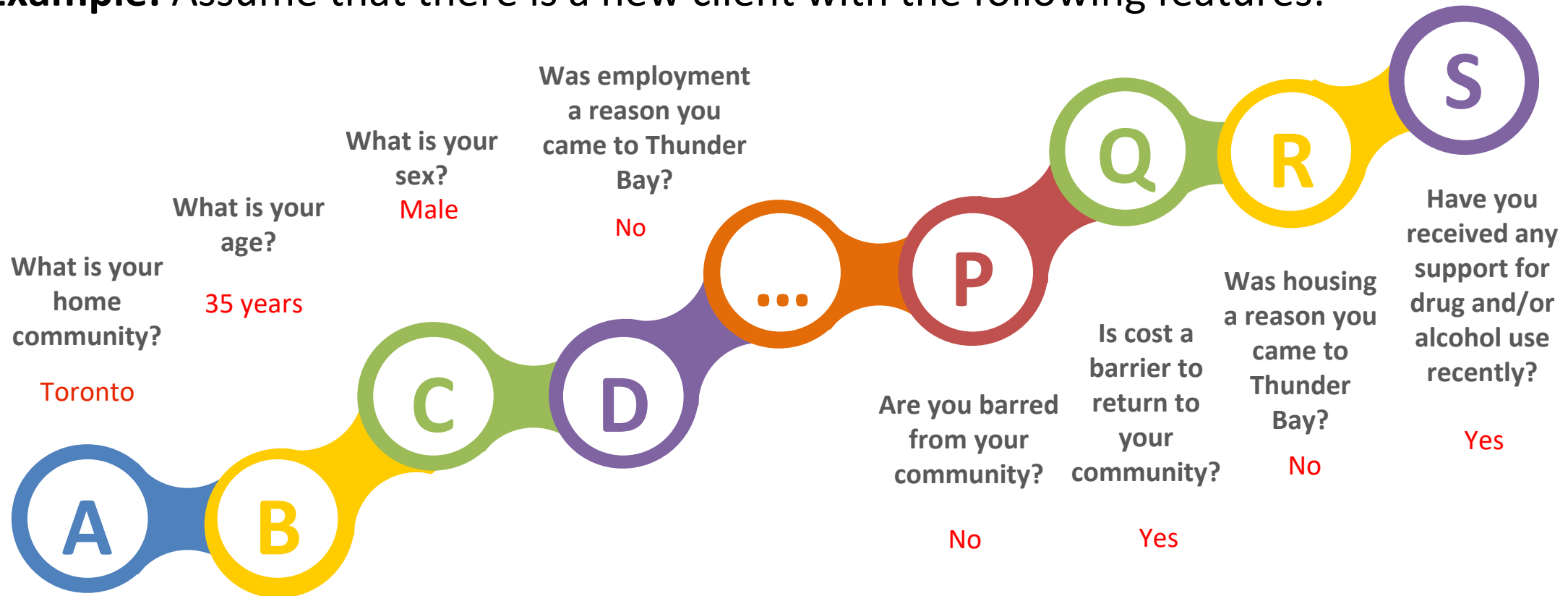# Results: RQ2 - If an individual stays, how long are they likely to stay?

- Performance of each model was tested with the best number of features and best synthetic data percentage.

- Logistic Regression performed well over time.

Performances of different models over time

Legend:
- Decision tree
- SVM
- Naïve Bayes
- KNN
- Random Forest
- Logistic Regression
- MLP
- Gradient Boosting Tree

Results: RQ2 - If an individual stays, how long are they likely to stay?

Example: Assume that there is a new client with the following features:

What is your home community?
Toronto

What is your age?
35 years

What is your sex?
Male

Was employment a reason you came to Thunder Bay?
No

Are you barred from your community?
No

Is cost a barrier to return to your community?
Yes

Was housing a reason you came to Thunder Bay?
No

Have you received any support for drug and/or alcohol use recently?
Yes

The most accurate model predicted : 21-58 days as the stay duration

Results: RQ3 - What factors predict stay duration?

26th February
Data Points: 119

26th March
Data Points: 140

9th April
Data Points: 151

Best Predictors

- Is cost a barrier to return to your community?
- What is your age?
- Have you received any support for drug and/or alcohol use recently?
- Was support for drug and/or alcohol use a reason you came to Thunder Bay?
- Do you have a Status Card?

Best Predictors

- What is your age?
- What is your sex?
- Do you have band membership?
- Was support for drug and/or alcohol use a reason you came to Thunder Bay?
- Do you have a Status Card?
- Have you received any support for drug and/or alcohol use recently?

Best Predictors

- What is your ethnicity?
- Do you have band membership?
- What is your age?
- Do you have a Status Card?
- Was support for drug and/or alcohol use a reason you came to Thunder Bay?
- Have you received any support for drug and/or alcohol use recently?

LAKEHEAD UNIVERSITY

SSHRC CRSH

DATALab

Discussion: Key Findings

- The important **factors predicting the stay or leaves** of an individual were found to be:
    - Have you been hospitalized in Thunder Bay?
    - Do you have family or friends in Thunder Bay?
    - Was your family or were your friends a reason you came to Thunder Bay?

- The important **factors predicting the stay duration** of an Individual were found to be:
    - What is your age?
    - Was support for drug and/or alcohol use a reason you came to Thunder Bay?
    - Have you received any support for drug and/or alcohol use recently?
    - Do you have a Status Card?

SSHRC CRSH

Lakehead
UNIVERSITY

DATALab

Discussion: Limitations

- Data scarcity
- Stability of results with more data
- Cross-sectional (snapshot) vs. Longitudinal study

Discussion: Implications & Next Steps

❑ Analysis of the interview data (qualitative analysis)
❑ TBDSSAB to use forthcoming report in policy discussions

THANKS!

Appendix

Survey Questions

- What is your home community?
- What is your previous community?
- What is your age?
- What is your sex?
- What is your ethnicity?
- What is the highest level of education you have completed?
- Do you have a Status Card?
- Do you have band membership?
- Was education/school a reason you came to Thunder Bay?
- Do you have family or friends in Thunder Bay?
- Was your family or were your friends a reason you came to Thunder Bay?
- Have you been incarcerated or to jail?
- Was court a reason you came to Thunder Bay?
- Did a fire, flood, or evacuation crisis force you to leave your home community?
- Was a medical appointment for yourself or a family member a reason you came to Thunder Bay?
- Have you been hospitalized in Thunder Bay?

Survey Questions

- Was employment a reason you came to Thunder Bay?
- Did you find employment when you arrived in Thunder Bay?
- Are you currently employed?
- Was housing a reason you came to Thunder Bay?
- Was mental health support a reason you came to Thunder Bay?
- Have you received any mental health support recently?
- Was support for drug and/or alcohol use a reason you came to Thunder Bay?
- Have you received any support for drug and/or alcohol use recently?
- Are you barred from your community?
- Is cost a barrier to return to your community?
- Do you feel that you can't go back to your community for personal reasons?
- Do you feel you can't go back to your community for any other reasons?