NSGROW

Nova Scotia Growth Resource Optimization Workflow

January 29, 2024

Kanav Bhardwaj | B00884919 | kn947171@dal.ca

Abhiroop Yerramilli | B00869102 | ab845358@dal.ca

Yakov Fainshtein | B00850311 | yakov.fainshtein@dal.ca

Matthew Carl | B00864121 | mt483195@dal.ca

Ji Lu | B00618068 | ji.lu@dal.ca

Introduction

In Nova Scotia, as in many regions worldwide, effective resource allocation is crucial for promoting economic growth, fostering community development, and ensuring fiscal sustainability. Municipalities face diverse challenges and opportunities; each requiring tailored strategies and investments to address local needs and capitalize on strengths. Poor resource allocation can result in problems such as stunted economic development, under/over investment, environment degradation, housing crisis, social unrest, to name a few.

Addressing these challenges and opportunities requires a data-driven approach to resource allocation that leverages insights from datasets such as the Municipal Fiscal Statistics. By analyzing trends, identifying patterns, and predicting future needs, policymakers and stakeholders can make informed decisions that maximize the impact of investments, promote sustainable development, and improve the well-being of all Nova Scotians.

Motivation

The proposed project offers a chance to contribute to the fiscal understanding of local governments in Nova Scotia. By examining five years of income statements, we can assess the financial operations and health of these entities. Our analysis aims to identify patterns in revenue and spending, informing more evidence-based policy discussions. In doing so, we hope to guide local governments towards improvements in spending effectiveness and service enhancement, supporting their journey toward fiscal responsibility. Our use of analytical techniques and predictive modeling is intended to refine the way numbers translate into fiscal strategy. This project is an exercise in diligence and prudence, with an eye towards reinforcing the foundations of accountable and transparent public service.

Benefits

This data science project aspires to offer meaningful benefits. It is an exercise in careful fiscal analysis, helping stakeholders identify areas for potential improvement in financial management within Nova Scotia's local governments. By systematically examining income statement data over a five-year span, we may uncover actionable insights into revenue optimization and prudent expenditure. The predictive models we plan to construct are intended as aids to better anticipate budgetary requirements, with the hope that public resources are judiciously directed where they are most needed. While the insights we derive could suggest adjustments to intergovernmental transfers and support, our overarching goal is to nudge policy toward paths that may lead to steady and responsible community growth. This initiative, grounded in academic rigorous seeks to be a contributor to the enhancement of local governance.

Description Of Dataset

The dataset for the proposed project is "Municipal Fiscal Statistics: Operating Fund Total Revenues and Expenditures by Municipality" The dataset is publicly available and presents the income statements for municipalities in Nova Scotia (from 2013 to 2020 fiscal years). This dataset includes the Non-consolidated Operating Fund Revenues and Expenditures for all the municipalities.

Data contains 26 features (dimensionality) and has a size of 408 data points.

Feature	Description	Type
Region	The area where data is gathered.	Categorical Nominal
Region Type	Type of region (e.g., municipality, rural).	Categorical Nominal
Year	The year in which the data was gathered.	Interval scale (treated as continuous)
Net Property Taxes and Payments in Lieu of Taxes	Total monetary value of property taxes paid.	Continuous Numerical
Assessable Property Taxes	Property taxes assessed, paid by personal or privately owned properties.	Continuous Numerical
Grants in Lieu of Taxes	Amount paid by government-owned buildings instead of taxes.	Continuous Numerical
Services Provided to Other Governments	Monetary services provided to other governments.	Continuous Numerical
Sales of Service	Fees paid to the government for utilities and other services.	Continuous Numerical
Other Revenues from Own Sources	Revenues from other government operations.	Continuous Numerical
Unconditional Transfers from Other Governments	Unwarranted transfers from other governments, often referred to as 'generosity transfers'.	Continuous Numerical
Conditional Transfers from Other Governments	Warranted transfers from other governments with specific purposes or conditions.	Continuous Numerical
Conditional Transfers from Federal or Provincial Governments or Agencies	Conditional transfers from larger government bodies.	Continuous Numerical
Conditional Transfers from Other Local Governments	Conditional transfers specifically from local governments.	Continuous Numerical

Total Revenues	The sum of all the above values representing any cash inflow to the government from that county.	Continuous Numerical
General Government Services	Services generally provided by the government, such as powering and installing streetlights.	Continuous Numerical
Protective Services	Services provided by the government to protect the population, including police, firefighters, paramedics.	Continuous Numerical
Transportation Services	Government-provided transportation services, which could include buses, roadwork, subways, etc.	Continuous Numerical
Environmental Health Services	Services provided by the government to protect the environment, such as park rangers and park cleanup.	Continuous Numerical
Public Health Services	Public health services, including paying doctors and nurses.	Continuous Numerical
Environmental development services	Environmental development involves the development of the environment within the region.	Continuous Numerical
Recreation and cultural services	Recreational and culture services could include construction of parks, or organization of a public event.	Continuous Numerical
Extraordinary or special items	Any costs that do not qualify under any of the other features.	Continuous Numerical
Total Expenses	The total amount of expenses needed to run the county.	Continuous Numerical
Operating Surplus	Surplus prior to financing and transfers.	Continuous Numerical
Total Financing and Transfers	Any financing or transfers that were needed to cover expenses.	Continuous Numerical
Annual Surplus	Total Final Surplus.	Continuous Numerical

Our Problem

A government agency has tasked our team with leveraging datasets to develop a predictive model aimed at optimizing resource allocation to municipalities in order to foster sustainable economic growth while minimizing fiscal inefficiencies. The objective is to determine the optimal distribution of funds to each municipality to maximize indicators such as GDP growth, housing development, and population expansion, while ensuring that surplus/deficit remains balanced and conducive to continued development in the respective localities.

Approach

We will use predictive modelling to infer the optimal resource allocations based on old data. More specifically, we will be using Linear Regression over a given set of features.

Data Preprocessing

There are missing values and outliers in the data which we need to preprocess before training our model. Furthermore, there are some fields with string data which we need to convert to numerical (discrete) before training our model.

Feature Correlation

We will analyze feature correlations to identify relationships between different variables in the dataset. This will involve calculating correlation coefficients and visualizing the relationships using techniques such as scatter plots or heatmaps.

Visualization

We will create interactive visualizations to present model outputs and insights to the client in a user-friendly and easy to interpret form.

Results

The results will be an optimal resource allocation table in a human-readable format such as CSV. Visualizations will help aid the client in making sense of the results.

Functionality

The functionality is divided into following sections:

Minimum Functionality

The minimum functionality is to create a predictive model that produces an optimized allocation of resources for different localities in Nova Scotia.

- Data Integration: Import and preprocess the selected datasets containing financial, housing, population, and environmental data for each municipality in NS.
- Feature Selection: Identify and select relevant features such as total revenue, total expenses, housing statistics, population demographics, and environmental risk factors.
- Model Development: Build a baseline predictive model to allocate resources to municipalities based on selected features.
- Allocation Optimization: Optimize resource allocation to maximize economic growth indicators while ensuring a balanced surplus/deficit for each municipality.

Expected Functionality

Expected functionality is as follows:

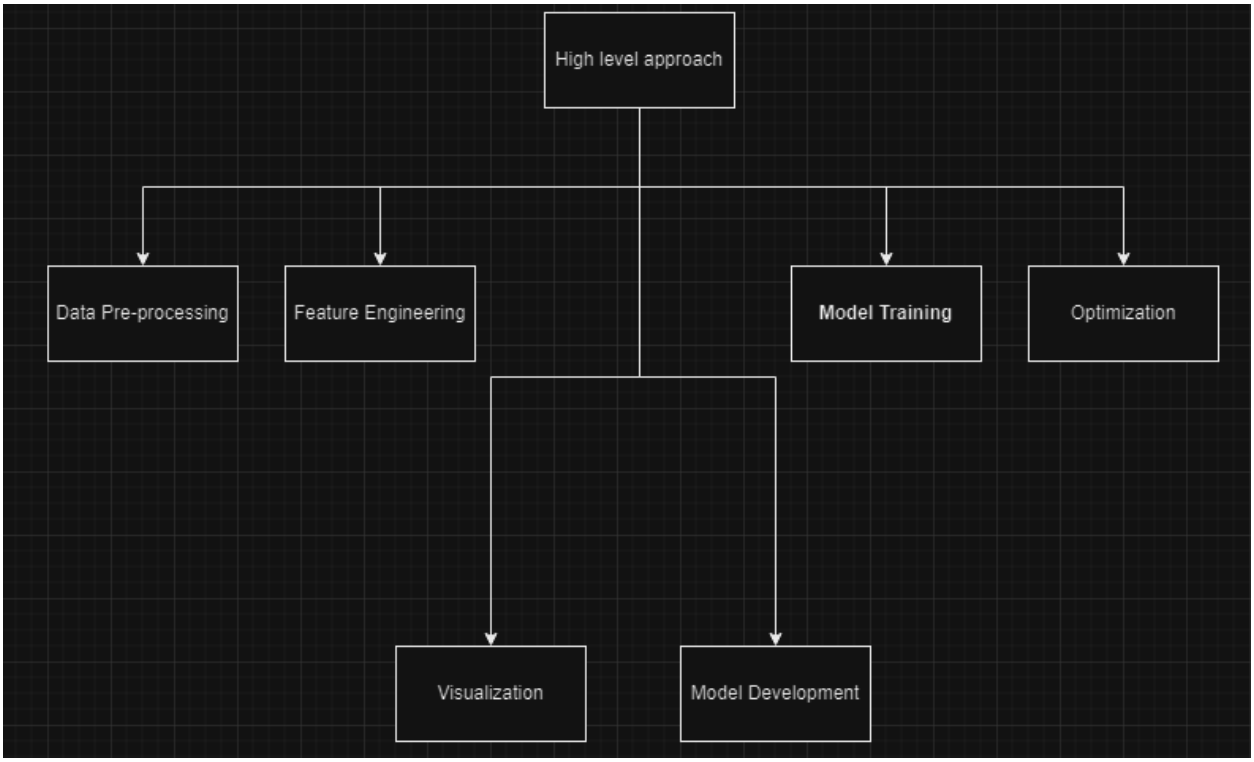
- Advanced Model Development: Implement more sophisticated machine learning algorithms to improve prediction accuracy and robustness.
- Constraint Optimization: Incorporate constraints into the model to ensure that surplus/deficit remains within predefined bounds while allowing for development.
- Validation and Testing: Validate the model using historical data and conduct rigorous testing to assess its performance and reliability.
- Temporal Adjustments: Consider the allocations from previous years when generating new allocations.
- Visualization: Develop interactive visualizations and dashboards to present model outputs and insights to client and/or stakeholders in a user-friendly manner.

Bonus Functionality

A user-friendly interface with following functionalities:

- Scenario Analysis: Enable users to conduct scenario analysis by adjusting input parameters (e.g., economic growth projections, housing policies) to explore different resource allocation strategies and their potential impact.
- Geospatial Analysis: Integrate geospatial analysis capabilities to visualize resource allocation decisions on maps and identify spatial patterns and disparities.
- Real-time Updates: Implement a mechanism to update the model with new data in real-time, allowing for dynamic adjustments to resource allocation decisions based on changing economic, demographic, and environmental conditions.
- Machine Learning Interpretability: Incorporate techniques for model interpretability to explain the rationale behind resource allocation decisions and enhance transparency and accountability.

High Level Organization and Design



The high-level organization of NSGROW is structured around a modular, sequential framework that guides the process from raw data to actionable insights. Initially, the Data Integration Module lays the foundation by aggregating, cleansing, and preprocessing financial, housing, population, and environmental datasets from various municipalities, ensuring the data's readiness for analysis. Following this, the Feature Selection Module takes over to meticulously identify and select the most impactful variables, such as total revenue, expenses, housing statistics, population demographics, and environmental risk factors, crucial for the predictive accuracy of the model. The core of the application is the Model Development Module, where a baseline predictive model is constructed using the chosen features to estimate optimal resource allocations, setting the stage for more advanced machine learning techniques that enhance prediction robustness and accuracy in the Advanced Model Development phase. To ensure the practical applicability of these predictions, the Allocation Optimization Module employs sophisticated optimization algorithms, focusing on maximizing economic growth while maintaining balanced municipal budgets, incorporating constraints to keep surplus/deficit within acceptable limits.

Milestones and Timeline

The following are the milestones we have set on a weekly basis:

Milestone 1 **(Week of Feb 2):** Data Cleaning and Preprocessing

- Examine dataset, identify errors, Implement initial data cleaning.
- Ensure data quality by addressing missing values and outliers.
- Standardize data formats and conduct transformations.
- Validate final dataset, ensure readiness.

Milestone 2 **(Week of Feb 9):** Descriptive Statistical Analysis

- Calculating summary statistics, and central tendencies.
- Analyze data distributions, identify key features.
- Explore correlations, assess data variability.
- Validate statistical analyses, ensure accuracy.

Milestone 3 **(Week of Feb 16):** Exploratory Data Analysis (EDA)

- Use visualizations to uncover patterns and trends.
- Identify outliers or anomalies through visual analysis.
- Explore relationships between variables and get insights.
- Validate findings, help with additional pattern understanding.

Milestone 4 **(Week of Feb 23):** Modelling

- Develop statistical models for the business problem.
- Validate outputs, testing, and assess suitability for the problem.

Week of March 1: Buffer (To catch up or start next Milestone early)

Milestone 5 **(Week of March 8):** Analysis and Realtime Updates

- Geospatial and Scenario Analysis implementation.
- Implement a mechanism to update the model with new data in real-time, allowing for dynamic adjustments to resource allocation decisions.

Milestone 6 **(Week of March 15):** Visualization and Reporting

- Interpret outputs of statistical analyses.
- Contextualize findings within the project's objectives.
- Create engaging dashboards or reports. Data presentation and storytelling (contextualizing with the business problem).
- Validate cohesiveness of visuals and the report.

Milestone 7 (**Week of March 22**): Policy Recommendations and Buffer

- Use key insights to make policy recommendations.
- Finish any pending work.