Investigating Anomalies in Energy Data:
A Case Study on Five Utilities in Ontario using Green Button and EBT Data

Anjali Mudaliar and Francis Palma
{anjali.mudaliar, francis.palma}@unb.ca
SE+AI Research Lab, Faculty of Computer Science, University of New Brunswick, NB, Canada.

Abstract

This paper analyzes and detects anomalies between Green Button (GB) and Electronic Business Transaction (EBT) data from Ontario utility providers. The study involved downloading and preprocessing energy consumption data from multiple meters, linking various XML files from each system, and developing Python scripts to automate the extraction process. A key analysis aspect was comparing GB and EBT data, particularly for Essex Power, an Ontario, Canada electricity utility, over one year. A One-Class Support Vector Machine (SVM) model was used to detect anomalies in energy consumption data. Our analyses revealed significant discrepancies between GB and EBT values for the same electric meter, with GB values consistently higher potentially due to missing factors in the data. We evaluated and identified patterns for each utility, such as spikes and dips in interval readings, highlighting the impact of data recording practices on anomaly detection. Challenges included addressing inconsistencies in data availability and formatting across utilities. This study lays the groundwork for future research into improving data quality and consistency in energy consumption records, with significant implications for energy management, accurate GHG accounting, sustainability initiatives, and utility data standardization.

Keywords: Green Button, EBT, Anomaly Detection, One-Class SVM, Utility Data, Energy Consumption, Multivariate time series.

1. Introduction

The Government of Ontario has recently mandated that utility providers offer consumers digital access to their energy usage data, marking a significant shift in how energy information is shared and utilized. This has led to the introduction of the Green Button (GB) initiative, which aims to provide easy access to energy usage data in a standardized format. By giving individuals direct access to their bill and meter data, Green Button provides information and empowers consumers to make informed decisions about their energy usage, contributing to broader decarbonization goals in the province.

Prior to this initiative, utilities in Ontario used the Electronic Business Transaction (EBT) standard for providing data to customers and retailers, managed by the Ontario Energy Board (OEB), which was designed primarily as a business-to-business (B2B) data exchange system. EBT was created to facilitate the secure and standardized energy data exchange between utility companies and their business partners, such as energy suppliers, without direct interaction with consumers. The EBT system requires that the parties that use the system have an Ontario Electricity Retailer License. The standard was limited to meter readings (start/end dates) and quantity consumed. While EBT effectively manages certain data exchanges between organizations, it lacks accessibility for individual consumers or other parties not “licensed energy retailers”. This limitation created a gap in energy data transparency. It hindered consumers from using the data themselves or innovators' help to make informed decisions about their energy consumption or to accurately measure the impact of energy management strategies on reducing consumption. The need for a more consumer-friendly system became apparent, leading to the implementation of the Green Button.

Green Button offers direct consumer access to all data fields on their energy bills and allows third-party energy management services to utilize this data via a standardized API. In contrast, EBT provides limited access, is restricted to retail license holders, and includes fewer data fields focused on business-to-business transactions. Green Button’s consumer-focused approach promotes innovation and enables tools like demand management and energy conservation, while EBT is more limited in scope and access. These technologies are crucial in reducing energy usage and ratepayer costs, promoting sustainability, and supporting Ontario's ambitious goals of reducing greenhouse gas emissions and transitioning to a low-carbon economy [1].

However, the introduction of Green Button data has raised new questions about the accuracy, consistency, and completeness of this newly available data. Because Green Button provides a different structure and data sets than EBT, comprehensive research is needed to evaluate data integrity, accuracy, and compatibility (with EBT). This is essential to ensure data meets the high standards required for reliable energy analysis and compliance with the "Best Available Data" policy by Ontario’s OEB. "Best Available Data" or BAD refers to utility data that is current, precise, and consistent, ensuring customers and third parties receive reliable information for decision-making, energy management, and regulatory compliance. For instance, if a utility provides energy consumption data but later adjusts it (due to recalculations or meter issues), it disrupts users' ability to rely on that data for real-time decision-making. Consistency in data delivery is crucial for customers and third parties to make informed decisions on energy management, regulatory compliance, and carbon reduction without the risk of sudden, unannounced changes.

Examples of how inaccurate or incomplete data could affect stakeholders:

Consumers: Incorrect energy usage data may lead to distrust in energy bills, making it difficult to track savings or identify ways to conserve energy. For residential consumers, errors in energy data can lead to mistrust in their bills and make it harder for them to track savings or find ways to cut down on energy costs. Energy managers depend on accurate data for businesses and institutions to plan budgets, optimize usage, and hit sustainability goals. When data is inconsistent or unreliable, these efforts complicate how they manage operations, control costs, and meet regulatory requirements. Both groups need precise data but for slightly different reasons.

Third-party innovators: Faulty data could disrupt demand management or energy conservation tools, hindering the ability to offer accurate solutions.

Regulatory bodies: Poor data quality could misinform policy decisions or skew conservation benchmarks, impacting Ontario’s decarbonization efforts.

Grid operators: Inconsistent data could impair the integration of Distributed Energy Resources (DERs), leading to unreliable grid demand forecasting.

This case study explores the techniques for data analysis and anomaly detection in energy data applied to both Green Button and EBT datasets. The research team evaluated seven meters with five utilities, evaluating the quality of data provided by each utility. By conducting this analysis, the research aims to identify potential discrepancies based on the evaluation categories between the two standards, providing insights into areas where Green Button data/EBT data may need improvement. The results of this research are expected to have far-reaching implications for Ontario and other jurisdictions globally, as many governments rely on accurate energy data for decision-making, carbon emission reporting, and implementing clean energy solutions. The results of this research are expected to have far-reaching implications for Ontario and other jurisdictions that have implemented or will implement the Green Button standard globally. Many governments rely on accurate energy data for decision-making, carbon emission reporting, and the execution of clean energy solutions. By ensuring data integrity, these jurisdictions can better support their decarbonization efforts and enhance the effectiveness of energy management strategies.

Various data analysis techniques were employed to conduct this research, including data preprocessing, cleaning, and normalization. Data normalization involves adjusting data values to fit within a standard range to ensure consistency in comparison, e.g., all energy readings are rescaled to a common unit. Python scripts were developed to automate data extraction and compare Green Button and EBT files provided in XML format. In addition, machine learning models, such as the One-Class Support Vector Machine (SVM), were utilized to detect anomalies in the energy data, helping to identify any irregularities or spikes in energy usage that could indicate issues with the data integrity. The analysis also considered the importance of time-series data, as both Green Button and EBT datasets contain multivariate time-series information that tracks energy consumption over time.

This case study also discusses the importance of data consistency across utilities. Data consistency means uniformity in recording data across all accounts and utilities. This includes consistent timestamps (e.g., billing data is always recorded at 11:59 pm), intervals (e.g., hourly readings maintained hourly across all records), billing cycles (e.g., starting on the 10th of each month), and units of measurement (e.g., always using kWh). Such consistency is crucial for accurate analysis and comparison. While Green Button (GB) provides a standardized data format, not all Ontario utilities fully adopt or implement it consistently. Some utilities still use legacy formats like EBT or customize GB data with non-standard fields or unique time intervals. With over 50 regulated energy utility providers operating in Ontario, each using different formats for their energy data, standardization is needed to ensure that data from various sources can be accurately compared and analyzed. The findings of this research highlight the key differences between Green Button and EBT data, emphasizing the importance of maintaining high data quality to support energy conservation efforts and the broader energy market.

The implications of this research go beyond Ontario. Many regions worldwide face similar challenges in transitioning to digital energy data portability. As governments worldwide seek to reduce carbon emissions and improve energy efficiency, reliable and consistent data is critical [2].

By understanding the strengths and weaknesses of both Green Button and EBT data, utility providers and government regulators can take steps to enhance the reliability of their customer's energy data. This will, in turn, benefit consumers by providing them with more accurate information to guide their energy decisions while also supporting the development of new energy-saving technologies that rely on high-quality data. The outcomes of this research can contribute to shaping energy policies, improving utility data practices, and advancing Ontario's progress toward a more sustainable, low-carbon future.

2. Background

2.1 EBT and OEB

The Electronic Business Transaction (EBT) system was developed in the early 2000s to improve business processes within Ontario's energy retail market. Initiated by the Ontario Energy Board (OEB), the EBT system is a digital platform allowing utilities and their business partners to automate various transactional exchanges. These transactions typically involve tasks like customer enrollments, third-party electric and natural gas billing, and meter data exchange, all essential for maintaining efficient utility operations with energy retailers across the province to settle energy transactions between multiple parties. The EBT system was designed specifically for business-to-business (B2B) interactions, facilitating communication between utilities and energy retailers rather than individual consumers. Historically, this system has played a vital role in Ontario's deregulated electricity market by allowing energy providers to electronically manage tasks like settling accounts and processing payments, reducing the need for manual intervention. This automation helps improve the accuracy and speed of data exchanges, making utility services more streamlined and reliable.

At the heart of the EBT system is the OEB, the regulatory body that ensures the system's standards are upheld. The OEB oversees energy transactions using the Retail Settlement Code (RSC) framework. This code outlines the rules for processing market transactions and handling customer data, ensuring fairness and consistency across all participants in the energy market.

One of the core components of the EBT system is the use of EBT hubs, as shown in Figure 1, which serve as centralized platforms where transactions are validated and routed between different regulated utilities and their OEB-licensed business partners. These hubs play a crucial role in ensuring that the transactions meet the standardized XML formats and protocols laid out by the EBT system. Additionally, they include error-checking functions that help catch issues before they disrupt the data flow. By providing this layer of oversight, EBT hubs help maintain the integrity and consistency of data exchanges across the entire energy market.

Although the EBT system has proven highly effective for facilitating B2B transactions, it was not created with consumers in mind, and the standard has not been improved over the years. This limitation became evident as the growing need for consumer access to their consumption data, improved security, and ease of access to digital data led to the development of the Green Button (GB) standard (3.3) ratified by the North American Energy Standards Board (NAESB). While EBT focuses on enabling secure data exchanges between businesses, Green Button aims to make energy data more accessible to individual consumers, allowing them to better manage their energy usage through innovative digital platforms.

The OEB is crucial in ensuring that the EBT system functions efficiently and that all market participants comply with its standards. As Ontario's energy market continues to evolve, the foundational role of the EBT system remains essential until a new infrastructure for the transactions is provided. As newer consumer-focused initiatives like GB roll out, it can be assumed that the older infrastructures will be retired. The introduction of Green Button marks a shift toward greater transparency, security, and accessibility, building on the structure provided by EBT.

2.2 Green Button Standard

The Green Button Standard was created to give consumers control over energy data in a standardized digital format. This initiative encourages the utilities to be transparent while helping consumers make informed decisions about their energy use [3], [4]. Utilities in Ontario are required to provide the best available information used "in the normal course of the energy provider's operations". Consumers are empowered to take charge of their energy habits and the authorization of where their data can be reused.

For consumers, this standard ensures that the information they access or authorize others to access is reflective of actual energy use and reliable for tracking consumption habits, budgeting, or participating in energy management programs. It emphasizes that utilities keep data up to date, promptly correct errors, and provide consistent quality across all accounts.

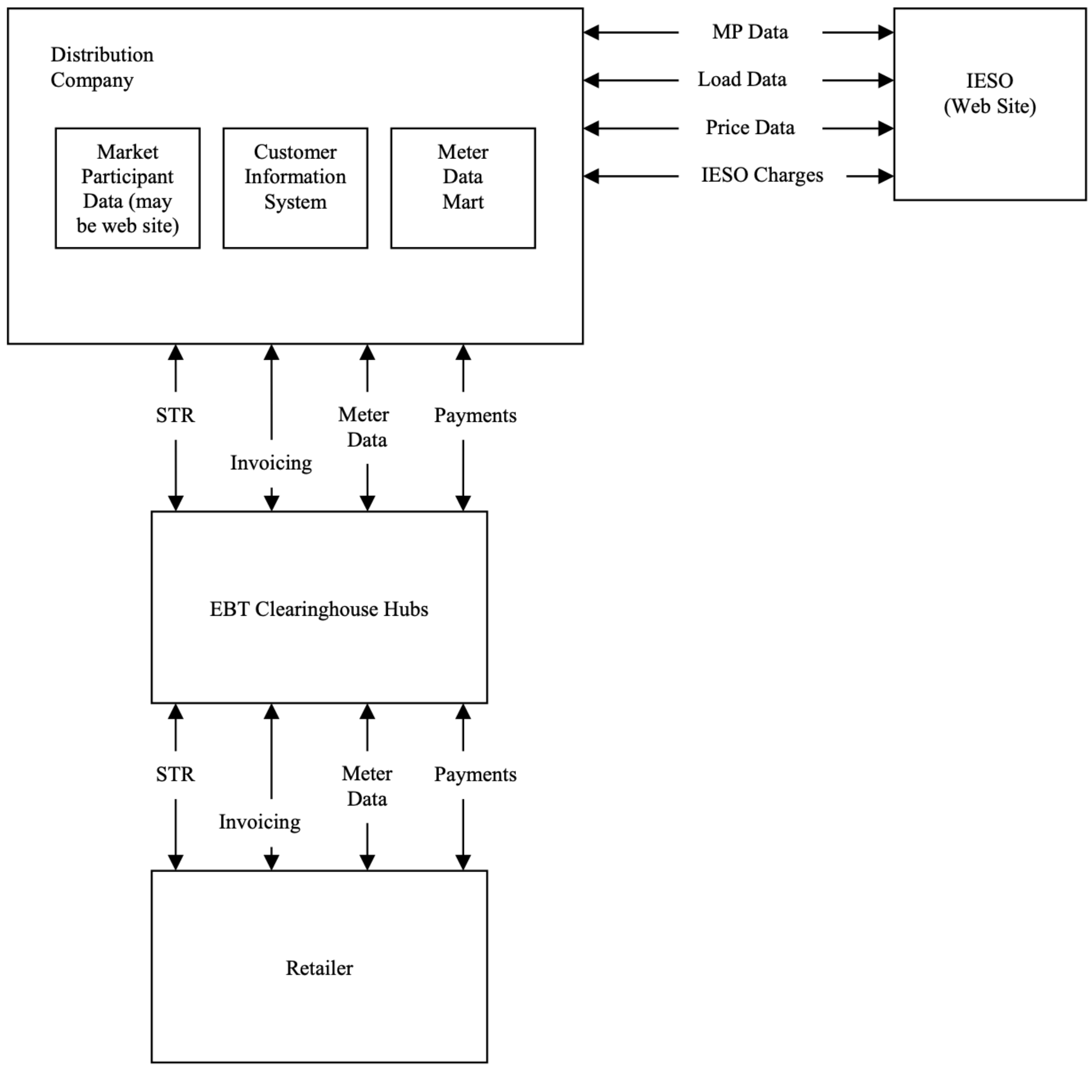

Figure 1: Electronic Business Transaction [11].

GB is designed with a consumer-to-business (C2B) approach, which differs from older systems like EBT. With Green Button, consumers can log into their utility's website or mobile app and transfer their energy usage data to any third party in a standardized XML format, as depicted in Figure 2. This data is updated at regular intervals, often daily, depending on the utility's systems. Once delivered, the third-party application analyzes it, helping consumers spot trends, find the most expensive times of day to use energy, or even get tips on reducing consumption or GHG output. For example, if a household notices their energy use spikes during certain hours, they could adjust their activities to reduce their bill.

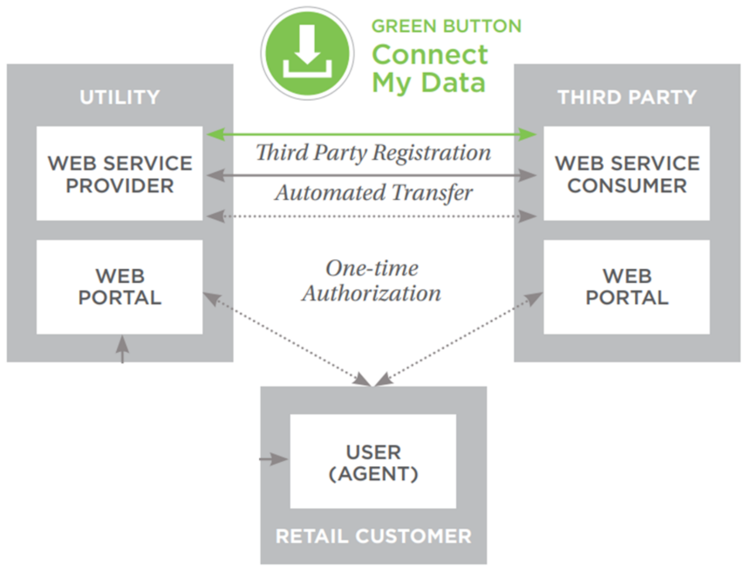

Figure 2: Green Button Connect (Credit: Michael Murray, Mission:data Coalition)

Green Button plays a pivotal role in promoting transparency in the energy sector. It encourages utilities to be more open about the energy data they collect, use, and manage for the customer. This transparency, driven by the OEB's policy for Utilities to provide "best available data," ensures that consumers are well-informed about their energy usage and how it is measured. To clarify the need for transparency, utilities' data is not always directly accessible or interpretable for consumers. The "Best Available Data" policy pushes utilities to provide energy usage data in a consumer-friendly format, offering insights beyond the basic meter reading. Without this, consumers may only see final billed amounts without knowing details like peak usage times, intervals, or adjustments (e.g., loss factors). Transparent access to comprehensive data lets consumers understand their consumption patterns more fully and empowers them to make informed energy management decisions. This builds trust in the marketplace and supports use cases such as energy-saving / GHG reduction programs or green generation initiatives such as solar.

Green Button's transparency not only makes energy usage data more accessible for consumers but also provides valuable insights for businesses and energy managers. By enabling easy analysis of real-time, accurate data, Green Button helps inform decisions about energy-saving investments and promotes changes in usage behavior. This data-driven approach supports broader environmental goals, including Ontario's carbon reduction efforts, by helping consumers and businesses optimize energy use. When multiplied across millions of households, this reduction can significantly impact overall energy consumption, reduce the need for new large generation facilities, and contribute to lowering greenhouse gas emissions.

2.3 Differences between GB and EBT

Although both involve energy data exchange, the EBT system and the GB standard serve different purposes. Their key differences lie in who they are designed for, what utility systems the data comes from, and how it is used.

EBT was built as a business-to-business (B2B) platform. It is designed to facilitate automated XML data exchanges between utilities and OEB-licensed businesses, like energy retailers and service providers. Its primary role is to support behind-the-scenes transactions such as retailer billing, historical meter data exchanges, settlement, and retailer customer enrollments. EBT is structured to handle the operational needs of a legacy energy market, ensuring that utilities can efficiently share necessary information with their licensed retailers securely and in a standardized format. However, EBT does not provide individual consumers with direct access to consumption data, and it does not provide additional utility bill details such as unit cost or demand electronically, as it focuses on transactions between a specific business sector rather than personal or portfolio energy management.

On the other hand, GB is a consumer-focused (C2B) infrastructure. Its primary goal is to empower consumers by giving them direct access to their energy data or authorizing its reuse to allow analysis to make informed energy management decisions.

In summary, EBT is a system implemented in 2002 for a legacy energy retailer industry. It is focused on exchanging data between companies, while GB is designed for consumers and third-party software firms. It aims to provide them with secured, automated access to energy data to encourage smarter, more efficient energy usage.

2.4 Roles and Responsibilities of Utility Providers

In Ontario, over 50 utility providers are vital in managing energy distribution. With the introduction of the GB standard, their responsibilities have expanded, particularly in terms of their commitment to ensuring access to accurate and timely electronic energy usage and utility bill data for consumers or others authorized by the utility account holder.

In addition to managing data, utilities must ensure that billed values match the actual meter readings and that data on consumer bills aligns with Green Button (GB) data for consistency. This includes verifying that usage calculations on the bill are accurate, transparent, and based on reliable data from the meter. Utilities are also responsible for protecting consumer data and securing digital data transfers to third parties, adhering to Ontario Energy Board (OEB) regulations to balance accessibility with security.

3 Study Method

This section outlines the detailed steps followed in data collection, preprocessing, and analysis to compare and detect anomalies between Green Button (GB) and Electronic Business Transaction (EBT) data from Ontario utility providers. The study collected meter data from the following utilities: Essex, Hydro One, Toronto Hydro, London Hydro, and Enbridge Gas. It also included model building for anomaly detection in the Green Button dataset for the above utilities.

3.1 Data Collection

Significant delays in receiving data from several utilities, both GB and EBT, were due to their internal system issues. This impacted the pace of analysis, as data collection was achieved many months after the market opened (November 2023), and Utility systems were still not thoroughly tested.

While this analysis took place more than six months after the compliance deadline, many Ontario utilities were still not fully aligned with Green Button (GB) standards. Additionally, the lack of rigorous testing by utilities, the Ontario Energy Board (OEB), and GB developers contributed to delays in implementation. These factors meant that utilities struggled to provide the required data in a compliant format, which affected the study. Data collection was delayed, limiting the pace of analysis and raising challenges around data consistency and reliability in this evolving system.

Green Button data was provided in XML format on UEnergyHub [5], a data research portal provided by Screaming Power Inc., consisting of customer resource files and energy usage files containing essential details such as Account IDs, Usage Points, and meter readings over time ("Interval Blocks" in GB parlance). Scripts were required for data extraction [6], and manual intervention was required to ensure correct linkage between bill and meter data because the customer resource and energy usage files were downloaded in pairs from the UEnergyHub portal without the utilities providing a link between the files. Due to the absence of a built-in link between the customer resource files and energy usage files provided by utilities, we had to develop custom scripts and manually link the data for accurate analysis. This lack of linkage on the UEnergyHub portal made data extraction and integration complex. The manual intervention required highlights a critical gap in how utilities provide data, adding significant effort and potential for errors, directly impacting the study's efficiency and accuracy.

In some cases, bulk data was received, as in the case of Hydro One and London Hydro, where separate scripts had to be written to handle the large volume of data. For instance, London Hydro's dataset was not internally linked, requiring manual intervention and the creation of customized scripts to preprocess and gather the data correctly. This challenge highlights a major barrier to seamless data integration, as the lack of bulk authorization protocols increased complexity and processing time.

The research period was limited to a few months, but we made progress in developing extraction scripts and identifying discrepancies between Green Button and EBT data. One of the key findings was the value of the Essex Power data. This was the only dataset with seven common accounts available in both GB and EBT, allowing for a comparative analysis of energy usage patterns. This unique opportunity provided valuable insights and underscored the importance of Essex Power data in our study. Table 1 below summarizes the number of accounts available in Green Button (GB) and Electronic Business Transaction (EBT) formats, common accounts between the two, and the interval periods for meter data collection across various utilities.

Utility Name	**Number of accounts available (GB)**	**Number of accounts available (EBT)**	**Common Accounts**	**Interval period for meter data (GB and EBT)**
Enbridge Gas	5	15	0	Monthly
Hydro One	3	2	0	Hourly
Toronto Hydro	3	2	1	Hourly
Essex Power	7	26	7	Hourly/Monthly (varies with account)
London Hydro	10	15	0	Every 15 minutes
Lakefront Utilities	1	0	0	Monthly

Table 1: Account Availability and Meter Data Intervals for Ontario Utilities

3.2 Data Preprocessing

Once the data was collected, significant preprocessing steps were necessary to clean and prepare the datasets for analysis. The preprocessing varied depending on the source and utility provider, as inconsistencies in format and missing values were the most common challenges.

Once the data was extracted into CSV files, several cleaning steps were necessary to ensure consistency and prepare the data for further analysis. Since billing data was only recorded once a month, while interval readings were captured at more frequent intervals (daily or hourly), fields such as reading quality and other billing-related information were often missing for the interval data.

A combination of forward-fill (ffill) and mean imputation was applied to handle these missing values. Forward-fill was used to propagate the most recent available billing information across the interval records. Where gaps persisted, the mean of the corresponding fields was applied to maintain data integrity and completeness.

The EBT and Green Button (GB) datasets had format inconsistencies for timestamp columns. The EBT timestamps were in the format YYYYMMDDHHMMES (e.g., 202405010000ES), while the GB timestamps were in epoch time (Epoch time, also known as Unix time, represents timestamps as the number of seconds (or milliseconds) that have passed since January 1, 1970, UTC. For example, 1640995200 in epoch format corresponds to 2022-01-01 00:00:00 in human-readable date format.). We converted both formats into a standard DateTime format, ensuring that the time data across both datasets was consistent. This standardization was crucial for accurate time-series analysis and comparison between the two datasets.

Another important data cleaning step in EBT data involved the energy reading units. The interval reading data used a mix of units, including kW, kWh, and kVA. To make the data consistent and comparable, all energy readings were converted to kWh, ensuring uniformity across both datasets for subsequent comparison and anomaly detection.

These data-cleaning steps were fundamental to ensuring the quality and consistency of the data for the subsequent stages of analysis and model building.

3.3 Data Analysis

The comparison between GB and EBT data for Essex focused on the seven common accounts. As in Figure 3, using a merger operation on the two datasets, the time periods for both were aligned, and the interval readings for the same accounts were plotted against one another. The results indicated that the GB values were significantly larger than the corresponding EBT values.

A logarithmic scale was used to visualize these discrepancies, as the difference between GB and EBT values was too vast to be plotted on a regular scale. A mean factor of difference of 965,717.04 was calculated, along with an average factor of 5,516,579.71, highlighting the substantial differences between the two datasets. After multiplying the EBT values by the mean factor, considerable overlap was observed, indicating that the missing Loss Factor in GB might contribute to the disparity. For example, if an EBT reading is 500 kWh and the Loss Factor is 1.05, the adjusted value becomes $500 \times 1.05 = 525$ kWh. Without this factor, GB's reading would appear lower than EBT's adjusted number, affecting direct comparisons. We observed overlap by scaling EBT values by the mean difference factor, but some inconsistencies remained, suggesting additional data anomalies. However, some values remained inconsistent after scaling, suggesting potential data anomalies.

BeginDate	EndDate	AccountID	UsageData (EBT)
2023-07-01	2023-07-31	Account1	9443.76
2023-08-01	2023-08-31	Account2	7455.60

IntervalReading_timePeriod_start	IntervalReading_timePeriod_end	AccountID	IntervalReading_value (GB)
2023-07-01	2023-07-31	Account1	1234
2023-08-01	2023-08-31	Account2	5687

BeginDate	EndDate	AccountID	UsageData (EBT)	IntervalReading_value (GB)
2023-07-01	2023-07-31	Account1	9443.76	1234
2023-08-01	2023-08-31	Account2	7455.60	5687

Figure 3: Merged EBT and GB Data using Inner Join for Account Comparison

3.4 Model Building

In this study, anomaly detection was performed using machine learning algorithms to identify irregularities in the GB energy usage data. After reviewing diverse options, we initially experimented with the Isolation Forest algorithm but selected One-Class Support Vector Machine (SVM) as the final model for anomaly detection.

We opted for One-Class SVM because it can handle multivariate time series data more effectively. One-class SVM works by creating a decision boundary around the normal data points and identifying any points outside this boundary as anomalies [7], [8]. This algorithm is particularly well-suited for detecting outliers in scenarios where the data points are related, such as energy readings over time.

Using scikit-learn [9], a popular Python machine learning library, we implemented One-Class SVM for anomaly detection. The model was trained on the GB data to capture typical energy usage patterns and then applied to identify data points deviating significantly from these norms.

Anomalies typically appeared as sudden spikes or drops in the energy readings that normal usage patterns could not explain.

The One-Class SVM successfully detected anomalies in the data, which were further verified through visualization techniques. We plotted time series graphs to highlight anomalies, showing where the model flagged sudden and unexpected changes in energy usage. This helped us pinpoint areas in the dataset where irregularities occurred, providing a clearer understanding of the data quality issues.

One-Class SVM proved to be the most effective approach for several reasons:

- Temporal Dependencies: It handled the temporal dependencies (a value depends on its past value) between interval readings and timestamps better than Isolation Forest, making it a more appropriate fit for time series data.
- One-class SVM's ability to simultaneously analyze multiple features, such as energy usage, timestamps, and reading quality, reassures us that it can find anomalies that simpler, feature-independent methods might have missed.
- Scalability: One-class SVM scales well for large datasets, which was important given the energy usage data we analyzed.

By leveraging One-Class SVM, we were able to detect anomalies efficiently and provide valuable insights into the dataset's irregularities.

4. Findings

This section summarizes the key findings of the study, categorized into overall data-related issues, utility provider-specific challenges, and observations from the comparison between the Green Button (GB) and Electronic Business Transaction (EBT) datasets. Each group highlights critical issues and offers recommendations based on the analysis conducted.

4.1 Overall Findings

4.1.1 Data Scale and Inconsistencies

***Issue*:** The most significant issue identified was the substantial difference in energy usage values between GB and EBT datasets. For example, in Essex, GB values were consistently higher than EBT values, with a mean difference factor of 965,717.04 (see Figure 4).

In Ontario, TOU pricing ranges from 8.7 cents per kWh during off-peak hours to 28.6 cents during peak times. However, many parties do not use TOU; they buy on the spot market, which can offer more flexibility and potential cost savings based on real-time prices. Each approach has its advantages depending on usage patterns. Applying the mean discrepancy factor of 965,717.04 between Green Button and EBT data suggests that the difference in usage could translate to a monetary impact of $83,019 to $276,197 per KWh, depending on the time-of-use rate. The absence of the loss factor is not enough to justify the difference, thus leading to the possibility of some more missing factors in the GB data. This large variance underlines the need for accurate loss and adjustment factors in Green Button data to prevent costly discrepancies. After applying the mean difference factor to EBT data (see Figure 5), many values aligned with GB data. However, some anomalies persisted, indicating potential data inaccuracies or missing factors in either dataset.

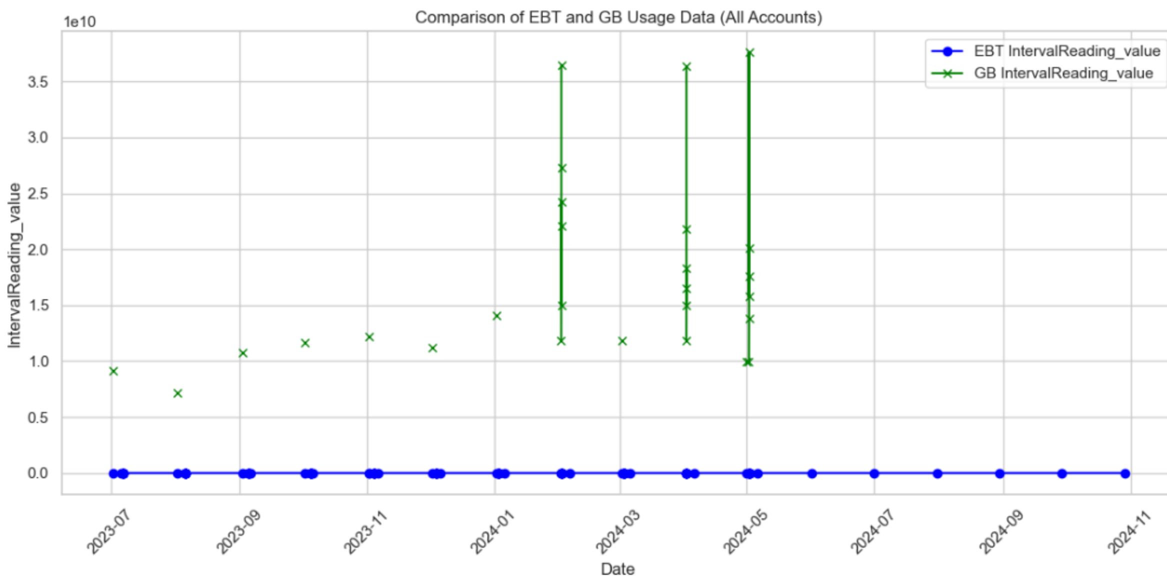

Figure 4: Comparison of EBT and GB Usage Data for all Accounts

Recommendation**:** Utilities should ensure that all relevant adjustment factors, such as loss factors, etc., are included in the GB dataset to provide more accurate and comparable energy readings as mandatory. Additionally, further investigation is needed into any remaining anomalies to determine whether they result from data inaccuracies or other underlying issues in the recording or reporting of energy usage. This recommendation should not be an issue to implement as the regulations and policies provided by the Government of Ontario already require this.

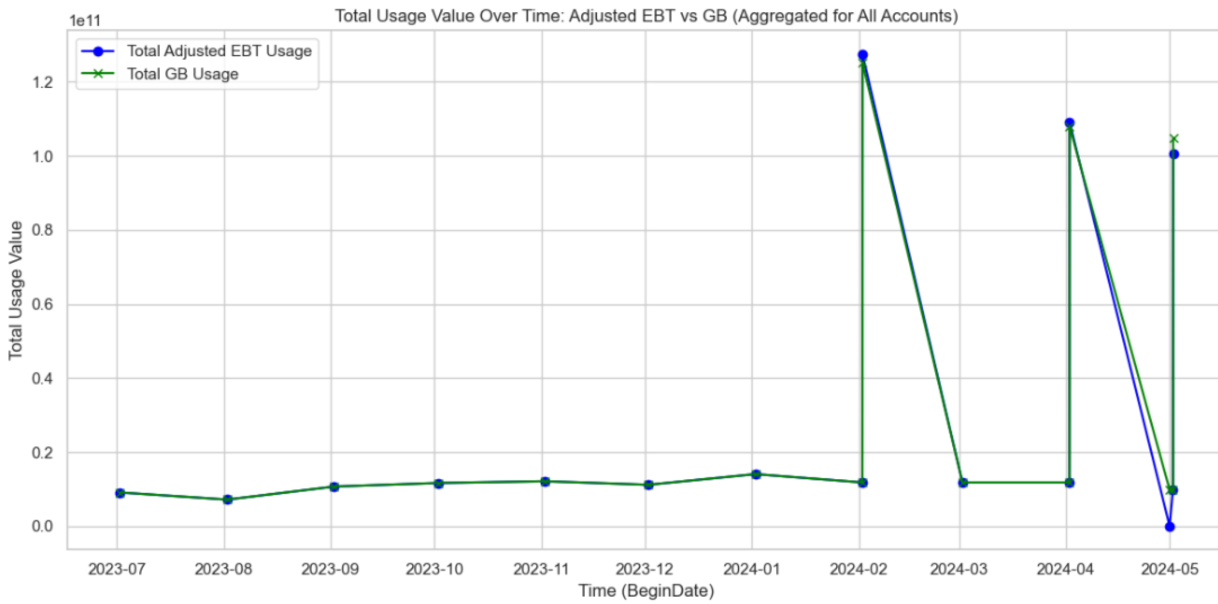

Figure 5: Total Usage Value After Adjusting Factor of Difference

4.1.2 Inconsistent Data Recording and Timestamps

***Issue*:** Another issue observed across multiple utilities was the inconsistency in how data was recorded in terms of usage unit, reading interval, or information provided. For example, in some accounts, energy readings were recorded daily; in others, they were recorded monthly, even within the same utility. This inconsistency made it difficult to analyze and compare data uniformly. Moreover, there were variations in the start and end times of interval periods, with some timestamps recorded as 11:59 and others as 00:00, further complicating the comparison and alignment of data between accounts.

***Recommendation*:** To improve data consistency, utilities should standardize the recording frequency across all accounts, ensuring that data is collected at consistent intervals (e.g., interval, daily, or monthly). This will significantly enhance the ease and accuracy of data analysis. Additionally, adopting uniform timestamps for interval periods will help align data and make it easier to analyze across different accounts and systems, further emphasizing the importance of data analysis in improving work efficiency.

4.2 Findings on Utility Providers

4.2.1 Issues from Toronto Hydro and Hydro One

***Issue*:** The Toronto Hydro dataset was more granular, with hourly readings, which allowed for more detailed analysis. Several unexplained spikes and drops were flagged as anomalies using the machine learning model for anomaly detection. Upon further investigation, one of these anomalies involved a residential customer with a sudden, unexpected spike in energy consumption over an hour, which was not typical.

In Figure 6, the y-axis represents the interval reading values of energy consumption recorded by Toronto Hydro, while the x-axis shows the time intervals at which these readings were recorded. The scale on the y-axis is in scientific notation, with values reaching up to 1.5e7, which equates to 15 million units (in watt-hours or a similar unit, depending on your data). This high scale captures both regular readings and anomalies effectively.

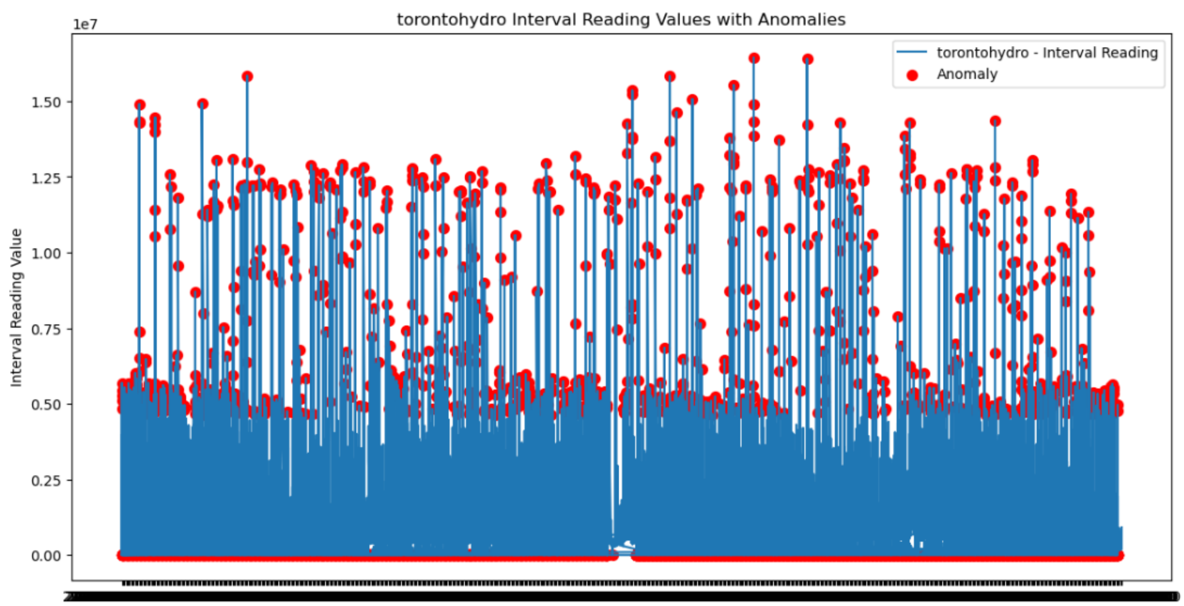

Figure 6: Anomalies in Toronto Hydro

The red dots highlight anomalies detected by the One-Class SVM model, representing unusual spikes or drops in consumption that deviate from typical patterns. These points are dispersed throughout the dataset, indicating fluctuations at various times. The continuous blue line represents the normal interval readings, providing a baseline for detecting these outliers. This visualization helps understand the intensity and timing of anomalies relative to standard energy usage, with the large y-axis range accommodating the significant variance in readings.

Recommendation**:** Toronto Hydro's hourly data makes it easier to detect and assess anomalies. Continuing with this level of data granularity will improve anomaly detection accuracy. Furthermore, investigating unexpected spikes in more detail can help identify underlying causes of anomalies, such as faulty meter readings or unusual energy usage patterns, potentially leading to more efficient energy management.

4.2.2 Issues from Enbridge Gas

***Issue*:** Enbridge Gas records energy data monthly, limiting the analysis' granularity. More frequent data collection (e.g., daily, hourly, or every 15 minutes) would provide better insights into energy consumption patterns and allow for a more accurate data quality assessment and implementation of energy management strategies. Additionally, Enbridge Gas provides bulk files with multiple customer data (for both resource files and customer files) in the same file. However, the customer resource file (containing customer details) is not properly linked to the resource file (containing meter data). Suppose there are 15 customers in the customer resource file with account details like Account ID, name, and address. Meanwhile, the resource file has meter data for these customers but lacks identifiers, such as Account ID or Usage Point, to match the data to specific customers. Without a common link, it is impossible to determine which meter data belongs to which customer, preventing accurate linkage between customer details and their energy usage. This requires manual or scripted intervention to establish connections. The reading cycle is also inconsistent, with some months starting on the 13th and others starting on the 15th.

***Recommendation*:** Enbridge Gas should move toward more frequent data recording (daily or hourly) to provide better granularity for analysis. Consistency in the recording cycle is also essential for reliable comparison over time. Moreover, linking the customer resource file with the resource file would allow the customer to use their files for easier management and analysis of energy data. This recommendation is especially important due to the number of customers this Utility provides for and the fact that they are the dominant provider of natural gas in Ontario.

4.2.3 Issues from Lakefront Utilities

***Issue*:** Like Enbridge and others, Lakefront Utilities records energy data monthly for the account used, which makes anomaly detection less effective due to the lack of granularity. In some cases, the reading cycle was inconsistent, with varying start dates from one month to the next, making it difficult to perform consistent time-series analysis. By increasing the frequency of data collection and ensuring a consistent recording cycle, Lakefront Utilities can significantly improve its anomaly detection capabilities, instilling confidence in the effectiveness of these changes.

***Recommendation*:** Lakefront should consider increasing the frequency of data collection to daily or hourly intervals to improve data quality and enable more detailed analysis. Ensuring a consistent recording cycle across months will also help accurately track energy usage patterns and detect anomalies. Note: only one account was provided. More review is needed to determine if this frequency is common in all accounts.

4.2.4 Issues from Hydro One

Issue**:** Hydro One provides hourly interval data, which is more granular than the monthly data from other utilities. However, the lack of common accounts between the EBT and GB datasets made direct comparisons difficult. Despite this, the hourly GB data appeared consistent and suitable for analysis.

Recommendation**:** Hydro One should make all accounts available via the Green Button to ensure accessibility and compatibility with current analysis methods. Additionally, implementing consistent hourly data collection for all accounts would enable comprehensive anomaly detection and data consistency checks across the utility. Bulk authorization for account access should be established to streamline data integration, allowing efficient, large-scale data retrieval for energy management and analysis.

4.2.4 Issues from London Hydro

Issue**:** The London Hydro dataset consisted of bulk files, with multiple customer accounts in the customer and resource files. However, only 10 of the 120 accounts in the customer resource file were linked to meter data in the resource file. This lack of linkage limited the scope of the analysis. Additionally, no common accounts between EBT and GB datasets prevented direct comparison.

Recommendation**:** London Hydro should ensure that all accounts in the customer resource file are properly linked to the corresponding resource file. Increasing the number of accounts available for analysis would also improve the utility's ability to compare datasets and assess data quality effectively. This change will also allow others to use the data.

4.2.2 Issues from Essex

Issue**:** Although seven common accounts were available for GB and EBT comparison for Essex, large energy usage value discrepancies were observed. Moreover, the data interval between GB and EBT was inconsistent. It was hourly for a few accounts and monthly for the rest. Therefore, aggregation of the data was needed for comparison. The missing loss factor and a few more data fields in Green Button data most likely cause these differences. After applying the mean difference factor, some values remained misaligned, suggesting data anomalies.

The EBT dataset provides a broad date range, with a beginning date of 2022 and an end date of 2024 across approximately 700 rows. However, there is no further breakdown to indicate specific days, months, or hours for each reading, making it unclear whether the data represents daily, hourly, or another interval. This lack of clarity complicates analysis and comparisons, as we cannot precisely match each reading to a specific date or time. Based on the number of records available, it was assumed that the recordings are hourly for comparison. The heatmap displays energy usage over time for various accounts in the EBT dataset, with darker shades representing higher usage values (see Figures 7 and 8).

Recommendation**:** For accurate energy consumption reporting, Essex should include loss factors and other components in its GB datasets. Further investigation into the anomalies should also be conducted to ensure data reliability.

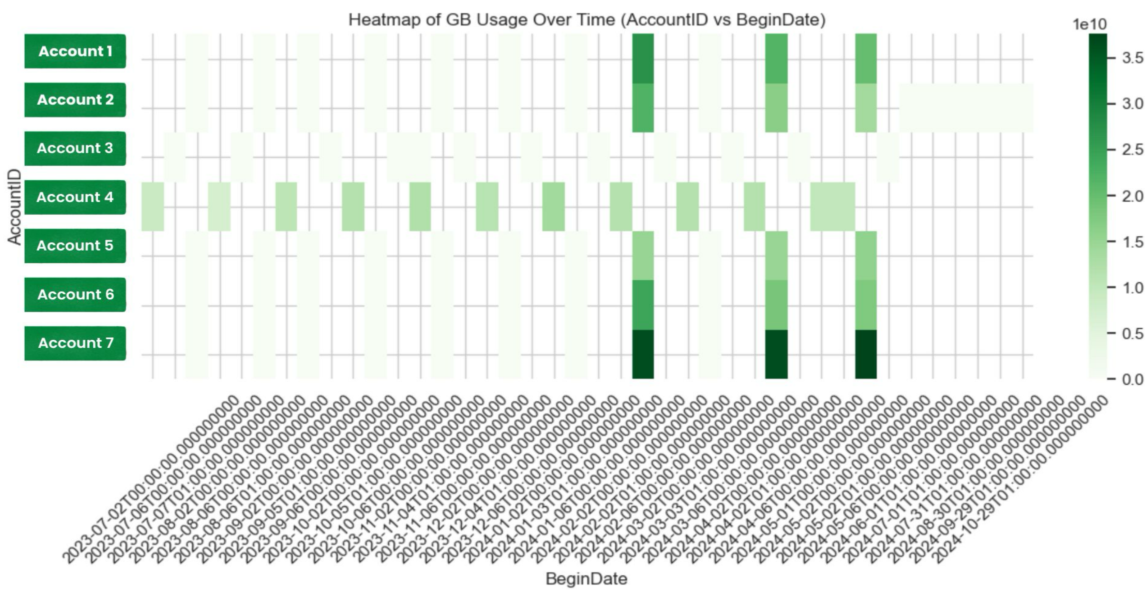

Figure 7: Heatmap of EBT Usage Over Time

5. ML-based Anomaly Detection

This section outlines the anomaly detection process using machine learning models, focusing on the input data, features selected, the target variable, and the analysis results.

5.1 Input Data

The input data for anomaly detection consisted of energy usage data from several utilities, including Toronto Hydro, Hydro One, Enbridge Gas, Essex, Lakefront Utilities, and London Hydro. Key data fields included utility name, account number, interval readings, quality of reading, and interval start and end times. Additionally, financial data, such as billing information, was used to flag anomalies like negative amounts. However, it did not directly affect meter readings and was not part of the model training. Preprocessing steps included removing duplicate values, filling missing data using mean and forward-fill (ffill) methods, and converting timestamps from epoch and EBT formats to a standardized datetime format.

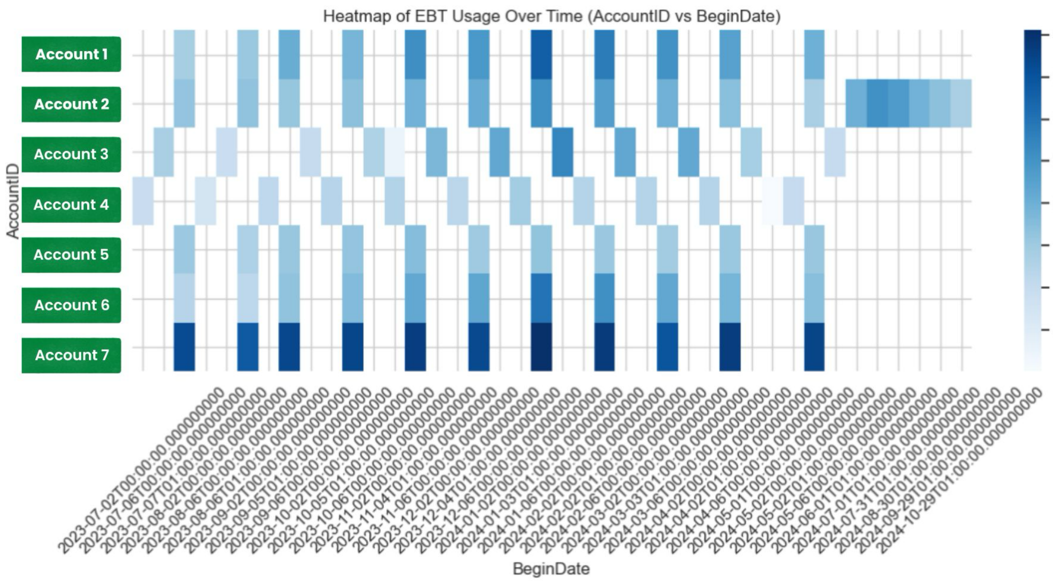

Figure 8: Heatmap of GB Usage Over Time

5.2 Features

The features selected for training the machine learning model were:

- Utility Name: Identified the utility provider used to segment data during analysis.
- Account Number: Unique identifier for each customer account.
- Interval Reading: The energy consumption values are recorded at specific intervals.
- Start and End (Interval): The timestamps for each interval reading are converted to a standard datetime format.
- Quality of Reading: A flag indicating whether the reading was valid, estimated, or manually adjusted [10].

These features were used to capture the energy consumption patterns and detect irregularities or anomalies in the data.

5.3 Target Variable

No explicit target variable existed since the model focused on unsupervised anomaly detection. Instead, using the abovementioned features, the One-Class SVM model was trained to recognize normal patterns in the interval readings. Any significant deviations from these patterns were flagged as anomalies.

5.4 Results

The One-Class SVM model successfully detected anomalies in the energy usage data, particularly for utilities with more granular readings, like Toronto Hydro, where hourly data allowed for identifying spikes and drops in consumption. The model flagged several unexplained anomalies, such as a significant spike in a residential customer's energy usage.

In contrast, utilities like Enbridge Gas and Lakefront Utilities, with monthly data, provided fewer anomaly detection opportunities due to the lack of detailed interval readings. Hydro One's hourly data appeared consistent, though no direct comparison with EBT data was made due to the absence of common accounts. Essex showed persistent anomalies even after scaling EBT data, suggesting potential data quality issues. The results were visualized using time-series graphs, showing clear spikes and drops in energy consumption across different accounts.

6. Recommendations for the Utility Providers

Based on our findings, we put together several recommendations for the utility providers as outlined below:

- Increase Data Granularity: Utilities such as Enbridge Gas and Lakefront Utilities should consider moving toward more frequent data collection (e.g., daily or hourly) to improve anomaly detection and assess data quality more effectively.
- Ensure Consistent Data Recording: Utilities should standardize their data collection processes. For example, interval start, and end times should remain consistent across accounts, and the recording cycle should be aligned month-to-month.
- Link Customer Resource and Meter Data: For utilities like London Hydro and Enbridge Gas, it is essential to link bulk customer resource files (with customer details) to meter data files to improve the scope and accuracy of analysis.
- Include Loss Factors in GB Data: Utilities should ensure that loss factors are included in GB datasets, as missing factors led to discrepancies in energy consumption reporting.
- Improve Timestamp Consistency: Standardizing timestamp formats across GB and EBT datasets would enhance comparability and make aligning and analyzing data easier.

7. Conclusion

In this project, we compared the energy data from both the EBT and GB systems to determine data integrity (completeness of data), data accuracy, and data compatibility/comparison between systems. To assess these, we applied One-Class SVM to detect anomalies in energy data from Ontario utilities, aiming to address inconsistencies and improve data quality.

We found that granular and hourly data provided rich insights, while monthly data, as seen with some utilities, limited anomaly detection details. This raised questions about why there was difference in the usage values for same account and interval periods in GB and EBT. Is the discrepancy merely due to missing factors like the loss factor in GB data, or are other adjustments needed? Without uniformity, the reliability of this data is compromised, which could have real-world impacts.

The inability of all utilities to implement bulk authorization creates significant inefficiencies in data management and access. Without bulk authorization, utilities cannot easily grant access to large sets of customer data, forcing data collectors to request access individually for each account. This manual process slows down data retrieval and increases the administrative burden on utilities and third-party analysts. The lack of bulk authorization limits timely access to aggregated data insights for customers, which can be critical for effective energy management and accurate billing analysis.

Comparing EBT and Green Button (GB) data highlights practical data integrity, accuracy, and compatibility challenges. EBT data, often tailored for business-to-business exchanges, can include loss factors and adjustments, making it more accurate for billing but complex for individual analysis. In contrast, GB data is designed for consumers, offering straightforward, real-time access but lacking these adjustments, which can lead to discrepancies when comparing the two. This mismatch impacts consumers and energy managers relying on consistency, as switching from EBT to GB may lead to differences in usage metrics, affecting energy management decisions and tracking.

Accurate utility data is foundational for all energy management, from driving informed economic decisions on energy efficiency, solar adoption, and other renewables to effective greenhouse gas (GHG) tracking to supporting Ontario's sustainability goals. How can consumers and policymakers make reliable choices if they cannot fully trust the underlying data?

Data analysis will inform Ontario's grid as it incorporates more distributed energy resources (DERs), and system operators manage demand dependent on consistent data. Utilities must adopt standardized data practices and tests to enable customers to plan better, manage, and forecast their energy consumption.

This project, while insightful, analyzed only a small subset of Ontario's 5.4 [12] million electricity and 3.8 million natural gas customers. For utilities to meet the Ontario Energy Board's "Best Available" data standard, compliance with Green Button must improve. Current gaps include limited bulk authorization, inconsistent implementation across utilities, and integrity issues when comparing Green Button (GB) and Electronic Business Transaction (EBT) data. Initially created in the U.S. to empower consumers, Green Button has potential, but further refinement is needed for widespread reliability and effectiveness.

Addressing these implementation failures allows Ontario utilities to enhance data transparency, support GHG reductions, ensure a more reliable energy grid, and strengthen trust in the energy market. This study highlights key findings in comparing Green Button (GB) and Electronic Business Transaction (EBT) data, emphasizing the need for improved data integrity, accuracy, and consistency across Ontario utilities. The analysis revealed discrepancies in data transparency, with GB lacking the loss factor present in EBT and limitations in access mechanisms like bulk authorization. Addressing these differences is critical for ensuring reliable energy data, enabling informed decision-making, and supporting Ontario's shift toward a more transparent and sustainable energy landscape.

8. Acknowledgement

We thank Mission:data Coalition and Screaming Power Inc for their financial support. We also thank the many industries, subject matter specialists, and utility account holders for providing their data through the UEnergyHub and EBT network.

9. References

[1] https://www.missiondata.io/news/2024/5/14/missiondata-announces-university-partnership-on-green-button-scorecard

[2] IESO. (2024). Ontario's Energy System Overview. Available at: https://www.ieso.ca/en

[3] Green Button Alliance. (2022). Green Button information. Available at: https://www.greenbuttonalliance.org

[4] Ministry of Energy: Regulatory proposal for province-wide implementation of Green Button, Last updated: September 16, 2021. Online: https://ero.ontario.ca/notice/019-2564

[5] Screaming Power Inc. (2024). *Energy Data Solutions*. Available at: https://www.screamingpower.ca/uenergyhub-empowering-ontario-for-a-greener-future-through-unbiased-research/

[6] The ElementTree XML API: https://docs.python.org/3/library/xml.etree.elementtree.html

[7] Sulaiman Kagumire: Anomaly Detection with Machine Learning, Available at: https://sulaimank.medium.com/anomaly-detection-with-machine-learning-8fa942fb5adc. May 12, 2021

[8] Scholkopf, B., et al. (1999). Support Vector Method for Novelty Detection. IEEE Transactions on Neural Networks, 12(3), 582-588. Available at: https://ieeexplore.ieee.org

[9] Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830. Available at: https://scikit-learn.org

[10] Green Button Utility API Document. Available at: https://utilityapi.com/docs/greenbutton/xml#SummaryMeasurement

[11] The Ontario EBT Working Group: Electronic Business Transactions (EBT) Standards Document for Retail Settlement in the Electric Retail Open Access Industry, Version 4.0, January 21, 2008. Online: www.oeb.ca/oeb/_Documents/Regulatory/Ontario_EBT_Standards_v4.pdf

[12] https://www.oeb.ca/sites/default/files/2024-04/Energy-at-a-glance-2022-2023-en.pdf