Sector-Wise Electricity Consumption Analysis at District level in Bangladesh Using Machine Learning: Patterns, Trends, and Predictions.

By
Zotirmoy Chowdhury
ID: 213-15-4306

Md. Tanvir Hossain
ID: 213-15-4582

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements forthe **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by
Mr. Raza Tariqul Hasan Tusher
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by
Md. Aynul Hasan Nahid
Lecturer
Department of Computer Science and EngineeringDaffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 16, 2025

APPROVAL

This Project titled **"Sector-Wise Electricity Consumption Analysis at District level in Bangladesh Using Machine Learning: Patterns, Trends, and Predictions"**, submitted by Zotirmoy Chowdhury, ID No: **213-15-4306** & Md. Tanvir Hossain, ID No: **213-15-4582** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **16 September, 2025**.

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Sadekur Rahman — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Saiful Islam (SI) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Arshad Ali (DAA) — **External Examiner**
Professor
Department of Computer Science and Engineering
Hajee Mohammad Danesh Science & Technology University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Raza Tariqul Hasan Tusher, Assistant Professor,** Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this projectnor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

14.9.25

Mr. Raza Tariqul Hasan Tusher
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by:

Md. Aynul Hasan Nahid
Lecturer
Department of Computer Science and EngineeringDaffodil International University

Submitted by:

Zotirmoy Chowdhury
Student ID: 213-15-4306
Department of Computer Science and EngineeringDaffodil International University

Tanvir

Md. Tanvir Hossain
Student ID: 213-15-4582
Department of Computer Science and EngineeringDaffodil University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Mr. Raza Tariqul Hasan Tusher**, **Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Big Data & Machine Learning** to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting themat all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

In this paper, I have examined electricity consumption patter and segmentation in Manikganj which has achieved full electrification but is still struggling with the issue of arrears. 'Manikganj Palli Bidyut Samity (MPBS) dataset' and various machine learning techniques, such as regression, classification, clustering, anomaly detection and causality, are applied in order to discover the patterns of electricity usage, consumer duties and support the efficiency and the awareness. XGBoost was the best in predicting (≈0.84) the variance of electricity but not the clusters hung high capacity low usage and low capacity high dues consumer. The Anomaly detection resurged an anomalous customer usage pattern which can indicate a suspective fraud or an inefficiency, whereas the causal effect proposed that the digital meter was associated with lower dues?.. as against analogue meter. Based on these findings, we have developed a web-based interface that envisages two views: The administration panel for back end support staff at the district to evaluate and regulate the data, and the user panel, where the customers can sign in to check their line capacity, the meter type, the usage report and payment clearance while a guest get access to the district-wide electricity information for educational learning and teaching as a learning resource. In sum, the studies and system exemplify the potential for data-driven intelligence and virtual tools to shed light on utility management and rural energy use and practices, make visible and transparent rural energy usage, and raise public awareness of both imaginary and actual rural energy use.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter describes the background and the motivation for the study. It sets the stage with a background of energy consumption worries in rural Bangladesh, explains the issue of arrears and inefficiency that the work is aimed at, and gives the motivation for undertaking such efforts. It also presents the goals to be achieved, the approach to be pursued, an estimated results list, and an outline of the report.

1.1 Introduction

Rural electrification has progressed a lot in Bangladesh, Manikganj being one of the first few districts to achieve full coverage. Despite this success however, the district yet has do deal with spin-offs from non-payment of electricity and those from revolving the outstandingness annually and their inevitable uncollectibility at the end of the financial year in June, which both militates against the performance of the district on both accounts, financial ad operational. These issues emphasize the need for understanding consumer behavior, detecting anomalous consumption and applying different methodologies for energy management. The aims of this research are to introduce the newest techniques of machine learning for sector-level electricity consumption patterns in Manikganj Division and seek usage behavior, comparative timely payers with denotative timely payers and defaulters, and predictive modeling in order to forecast future bills in the respect to Manikganj division. The intention is that these findings should improve billing reliability, remove loss making sections and present information for data driven decision making.

1.2 Motivation

The motives behind this project could be divided into several categories:

Computational motivation:

This is because of the fact that with increasing electricity consumption data, manual analysis becomes impossible (exponential growth of electrical read record volume). Machine learning is now available enabling scale-centric data-driven methods such as prediction, segmentation and anomaly detection.

Practical motivation:

It generates the consumption-based behaviors of defaulters and defilement patterns that can help electricity suppliers in designing an effective billing strategies, reducing some economic cost and delivering a progressing public alert

Motivation:

The authors see this work as an exercise to build a capability in data analytics and machine learning, and prepare their foundation before applying in practice in smart energy systems.

1.3 Objectives

The objectives of this project are clearly defined as follows:

- To investigate the sectorial pattern of electricity consumption in Manikganj.
- To determine the pattern of consumers and analyze their behavior using clustering and segmentation approaches.
- To forecast future power consumption with the use of regression models.
- To distinguish between Payable and Non-Payable Invoices to manage them better.
- To identify abnormalities and possible fraudulent use cases.
- To offer practical advice on billing integrity and awareness initiatives.

1.4 Methodology

These key issues of the research methodology are formulated as follows:

- **Data Collection and Preprocessing:** We have collected the dataset from Manikganj Palli Bidyut Samity (MPBS) and then we have preprocessed the dataset.
- **Exploratory Data Analysis (EDA):** 1 Applying statistical tests and visualization to uncover some relationships in energy consumption, due and there diffraction between the sectors.
- **Feature Engineering :** was able to create a complex feature such as usage per some KW capacity, risk scores of blocks in that sector.
- **Modeling:** Employ diverse machine learning models such as Classification (Logistic Regression, XGBoost, LightGBM), Regression (Random forest, XGBoost Regressor), Clustering (KMeans, DBSCAN) and Anomaly detection (Isolation Forest).
- **Performance:** Evaluation (Accuracy, Precision Recall, R^2 score, cluster visualization) of model.
- **Explanation:** Explainable model using SHAP values and interpretations through hypothesis testing.

1.5 Project Outcome

The expected outcomes of the research include:

- Industry-specific consumption of energy in detail.
- Commodities listing on the priority sectors that have a high frequency towards accumulation of dues.
- Predictive models for predicting future electric power consumption.
- Classifying people in different behavioral groups.
- Method for unusual usage pattern detection.
- Actionable intelligence for billing efficiencies and strategy around awareness campaigns.

1.6 Organization of the Report

The report is organized into the following chapters:

Chapter 1 — Introduction: Background, motivation, objectives, methodology, outcomes, and report organization.

Chapter 2 — Literature Review: Overview of related studies and identification of research gaps.

Chapter 3 — Methodology: Research design, dataset description, preprocessing, and model development.

Chapter 4 — Implementation and Results: Presentation of results, experimental findings, and visual analysis.

Chapter 5 — Engineering Standards and Design Challenges: Standards compliance, impacts, and design considerations.

Chapter 6 — Conclusion and Future Work: Summary of findings, identified limitations, and potential future research directions.

Chapter 2

Background

This section lays out the theoretical basis for the research and provides an overview of the key academic contributions in the field. It begins with a review of earlier studies on energy utilization and electricity distribution models, examining both their strengths and their shortcomings. It also considers comparable applications developed in other works or projects, highlighting the methodological approaches and results that are most relevant to the present study. In addition, this section incorporates a gap analysis to identify unanswered questions in prior research and to show how this project intends to address them through data-driven analysis and system development.

2.1 Introduction

Energy consumption studies have become an essential area of research for promoting sustainable energy management in developing nations such as Bangladesh. With rural regions—including districts like Manikganj—rapidly achieving electrification, electricity distribution has turned into a major issue that had little significance only a few years ago. Despite this progress, there is still no comprehensive framework for district-level monitoring, nor are there public awareness centers that track or analyze consumption trends in detail. Consequently, the use of electricity across different sectors and major commodities remains poorly documented, leaving policymakers and utility providers without sufficient insight into local energy dynamics. This lack of systematic monitoring has created a gap in understanding how energy is consumed, managed, and optimized at the rural level. To establish the foundation for this research, the next section presents a review of related literature, identifies the limitations of previous studies, and examines similar cases that help shape the direction and motivation of the present work.

2.2 Literature Review

This section will contain our literature review.

Table 2.1: Summary of Literature Reviewed.

Author (s)	Year	Title	Methodology	Key Findings
Joarder, Hakim & Mamun [1].	2010	Analysis of energy consumption and indicators of energy use in Bangladesh	Statistical and trend analysis	Identified key drivers of energy use in Bangladesh but lacked micro-level (district/rural) analysis.
Sarker, Farid & Alam [2]	2023	Analysis of the power sector in Bangladesh: current trends,	Policy analysis + data-driven study	Highlighted power sector challenges but no consumer-level or predictive

		challenges, and future perspectives		modeling insights.
Ahmed & Chowdhury [3]	2025	AI-Powered Energy Forecasting and Control for Smart Rural Energy Infrastructure	AI & ML-based forecasting	Showed potential of ML in rural energy but did not include awareness modules or district-specific insights.
Rahman et al. [4]	2025	Enhanced power demand forecasting for Bangladesh: using feature engineering	LSTM, Prophet, Random Forest.	Improved forecasting accuracy but focused on national demand, not sectoral/district segmentation.

2.2.1 Similar Applications

In Bangladesh, several electricity distribution companies—including Palli Bidyut Samity (PBS–REB), DESCO, NESCO, DPDC, and OJOPATICO—have already launched digital platforms to improve customer service, focusing mainly on online bill checking, payment processing, complaint management, and outage notifications. These initiatives have played a significant role in ensuring easier access and transparency for both rural and urban consumers through mobile financial services and web-based tools. However, the scope of these applications remains limited since they primarily serve transactional functions and do not incorporate advanced data-driven capabilities such as predictive forecasting of electricity consumption, consumer behavior segmentation, anomaly detection, or targeted awareness campaigns. In contrast to previous systems, the one presented in this paper is more than just a mere service system: it also brings bill tracking and consumer login functionalities of its predecessor assets, but gets intelligent for adaptive consumption profile per respective sector (machine learning), fraud detection and billing schedule. Besides, it has created a dual-channel platform, enabling authentic users to access their own consumption details and dues for each month at an iPhone touch, while guest users will view the statistics at district level in seconds. This pervasive approach guarantees the interworking of traditional billing systems and advanced energy management applications under a common information framework to supply consumers with customized tools, while administrators/decision makers could gain insight into data via analytics engines.

2.3 Gap Analysis

This table contain the gap analysis works between our work and other literature.

Table 2.2: Summary of gap analysis.

Features	**Joarder et al. (2010)**	**Sarker et al. (2023)**	**Rahman et al.**	**Ahmed & Chowdhury**	**Our Work**
Focus on District-level	No	No	No	No	Yes
Sector-wise Consumption Analysis	No	No	No	No	Yes
Due Payment Prediction	No	Yes	No	Yes	Yes
Clustering & Segmentation	Yes	Yes	Yes	Yes	Yes
Fraud / Anomaly Detection	No	No	No	Yes	Yes
Awareness-Oriented Web Application	No	Yes	Yes	Yes	Yes
Billing Simulation & Optimization	No	No	No	No	Yes

2.4 Summary

It is evident from the reviewed literature that most studies have concentrated on national-level analyses of the power sector, focusing on overall consumption trends or broad policy recommendations, while very little attention has been paid to district-level, sector-specific consumption patterns. This gap is particularly visible in rural and less developed areas such as Manikganj, where reliable data and systematic analysis remain scarce. An additional important consideration is that payment prediction, customer segmentation and profiling, fraud detection, billing optimization and knowledge based financial digital tools are identified as major analytics drivers, but are not detailed in incumbent studies. These challenges provide space for the present to be effective in that, it is able to provide insights on the real data and to design an intelligent system that can monitor the consumption of the sector, predict the user behavior and also to enhance the management efficiency and the awareness on the local level.

Chapter 3

Research Methodology

Here we present the method and study design. The content will act as the keyword in the case of approach and method the sector-wise consumption of electricity and non-functional/functional requirement of the project. DFD and UI design supported the system design. It then goes on to related work, why we did what we did and concludes with our project plan and summary.

3.1 Methodology

3.1.1 Overview:

In this paper, we employ both machine learning and expert knowledge domain to solve the task of analyzing energy consumption/ liabilities. Although the framework is not exclusively based on algorithmically created basis functions, our approach leverages concepts of learned features from datadependent samples and provides a way to ensure justified and interpretable results. Raw data was scraped & pre-processed from MPBS and is clean, encoded and organized for the analysis. An EDA was conducted to find patterns, relations between the sectors and usage distribution. Feature engineering was also applied to generate new feature sets, including a contrasting/relative set (opposed risk rating vs. capacity rating and an industry specific risk score) designed to enrich the model with added granular detail to draw actionable insights from the provided data sources.

3.1.2 Proposed Methodology/ System Design

The proposed methodology follows these steps:

Data Collection – Dataset collected from Manikganj Palli Bidyut Samity.
Preprocessing – Data cleaning, encoding categorical features (Type of Connection, Meter Type), handling missing values, and scaling numerical features (kW, kWh).
Exploratory Data Analysis (EDA) – Statistical and visual analysis of usage and dues across sectors.
Modeling –

- **Classification**: Logistic Regression, XGBoost, LightGBM, CatBoost for predicting dues.
- **Regression**: RandomForest Regressor, XGBoost Regressor for forecasting usage.
- **Clustering**: KMeans, DBSCAN, PCA for customer segmentation.
- **Anomaly Detection**: Isolation Forest, One-Class SVM for fraud/abnormal usage.

Evaluation – Metrics include Accuracy, Precision, Recall, R^2 Score, MAE, and visualization tools like confusion matrices and SHAP explainability.
Interpretation – Results analyzed for real-world implications in billing, awareness, and policy.

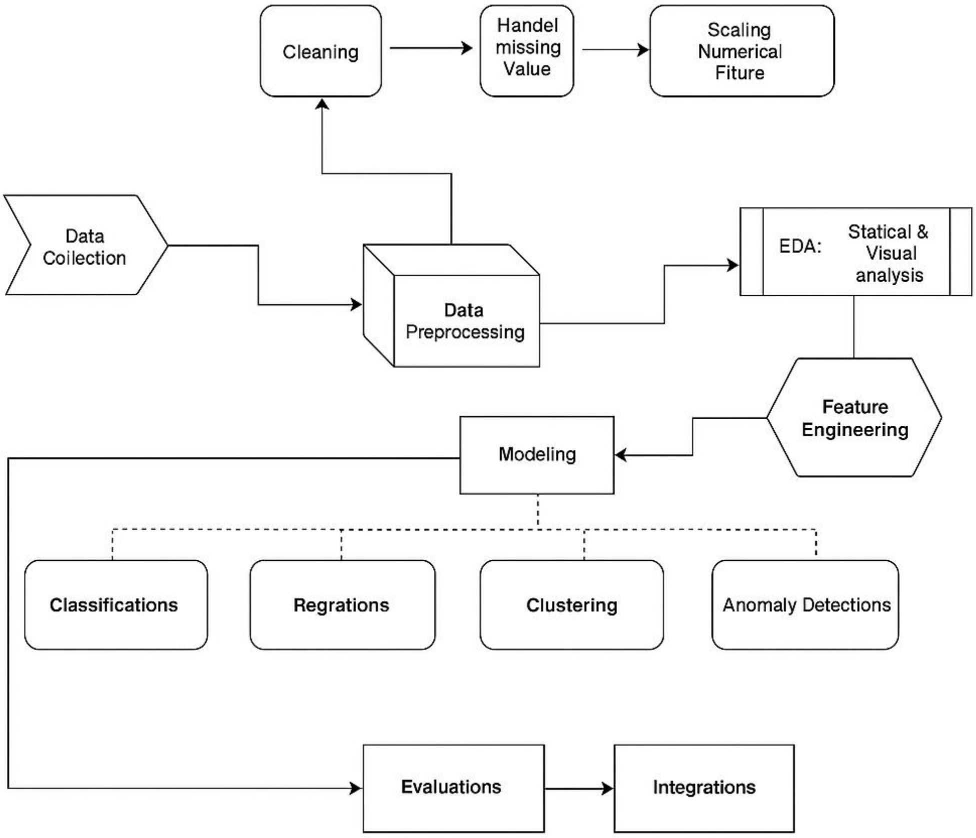


Figure 3.1: This is a sample diagram of our work.

3.1.3 Functional and Nonfunctional Requirements

Functional Requirements (What the system should do):

- Forecast the electricity usage of each sector.
- Classify customers as due and non-due.
- Classify customers by how and how much they use.
- Identify deviant consumption behaviors.
- Decision-making visual/dashboards.

Non-Functional Requirements (How the system should perform):

- Good prediction accuracy and robustness.
- Scalability for future datasets.
- Data Privacy and Security of consumer data.
- Shape/lime for interpretability / explainability.

Speed for scaling up predictions.

3.1.4 UI Design

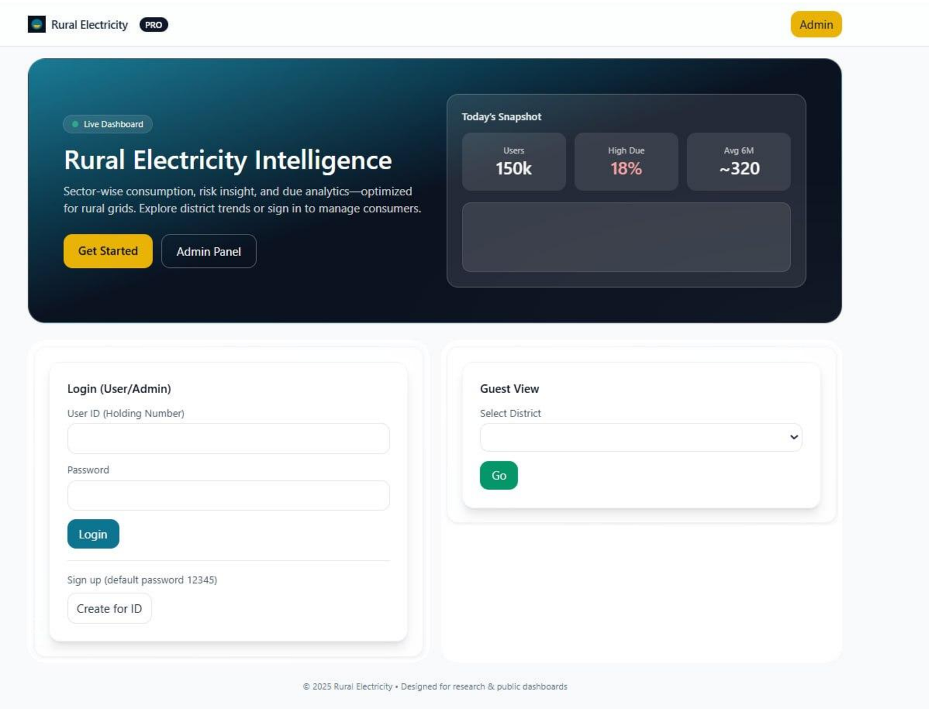


Figure 3.2: UI for everyone.

The project's campaign page points to a site only where you can glance at rural electricity. And there is an admin login form at the top. Some of the key features displayed on the main dashboard are real-time users snaps, due ratios are high and normal usage, etc. On the bottom login form is available for the users and admins and option to create new ID. Data can be viewed in guest view that is district wise too for non-registered users. All these make it easy for admins, registered users and guest.

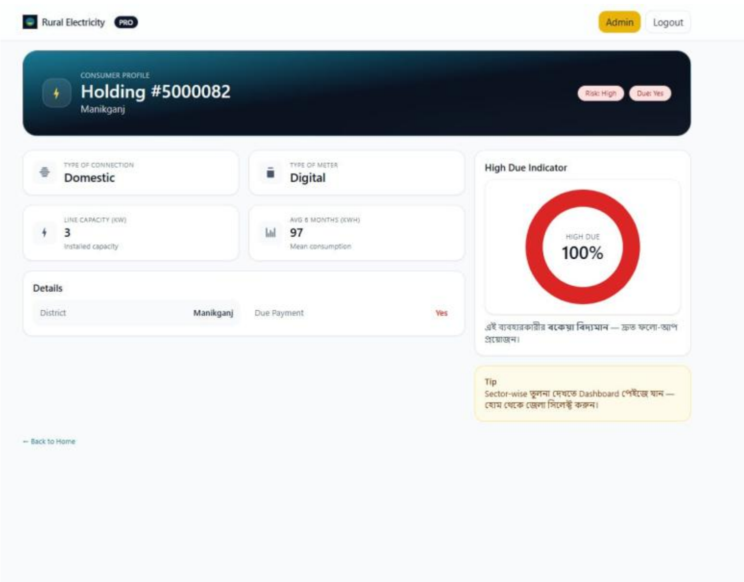
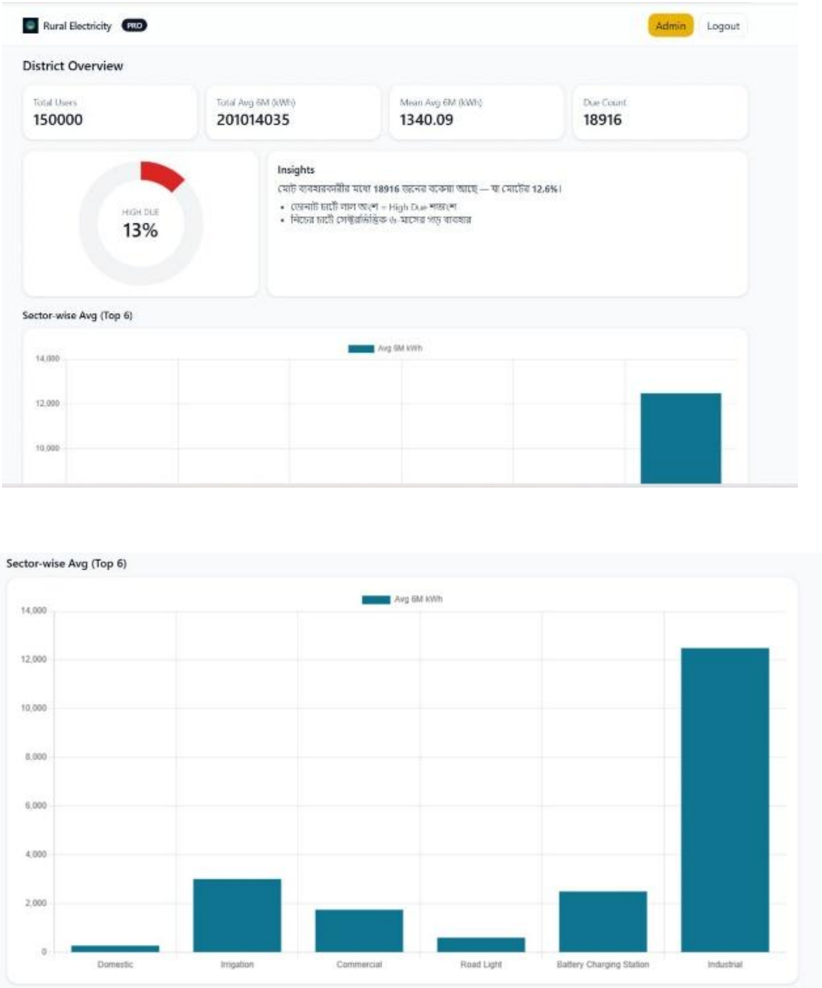


Figure 3.3: UI for Authentic Consumer.

Real users will need to log in with the username and passcode issued by the electricity office. When authorized users log in, they can view on screen the status and terms of their electricity connection.



Sector Table

Sector	Users	Avg 6M	Due
Domestic	89880	274.7	11696
Irrigation	22812	3006.3	6089
Commercial	14732	1748.6	743
Road Light	7616	602.0	63
Battery Charging Station	7371	2499.3	226
Industrial	4616	12487.1	41
Charitable Institution	2973	449.9	58

Figure 3.4: UI for the Guest User.

Visitors can obtain the gross electricity consumption as well as the gross customer consumption per district. The dropdown menu will display only such districts for which the uploaded data from their end (from the respective district electricity office) is available.

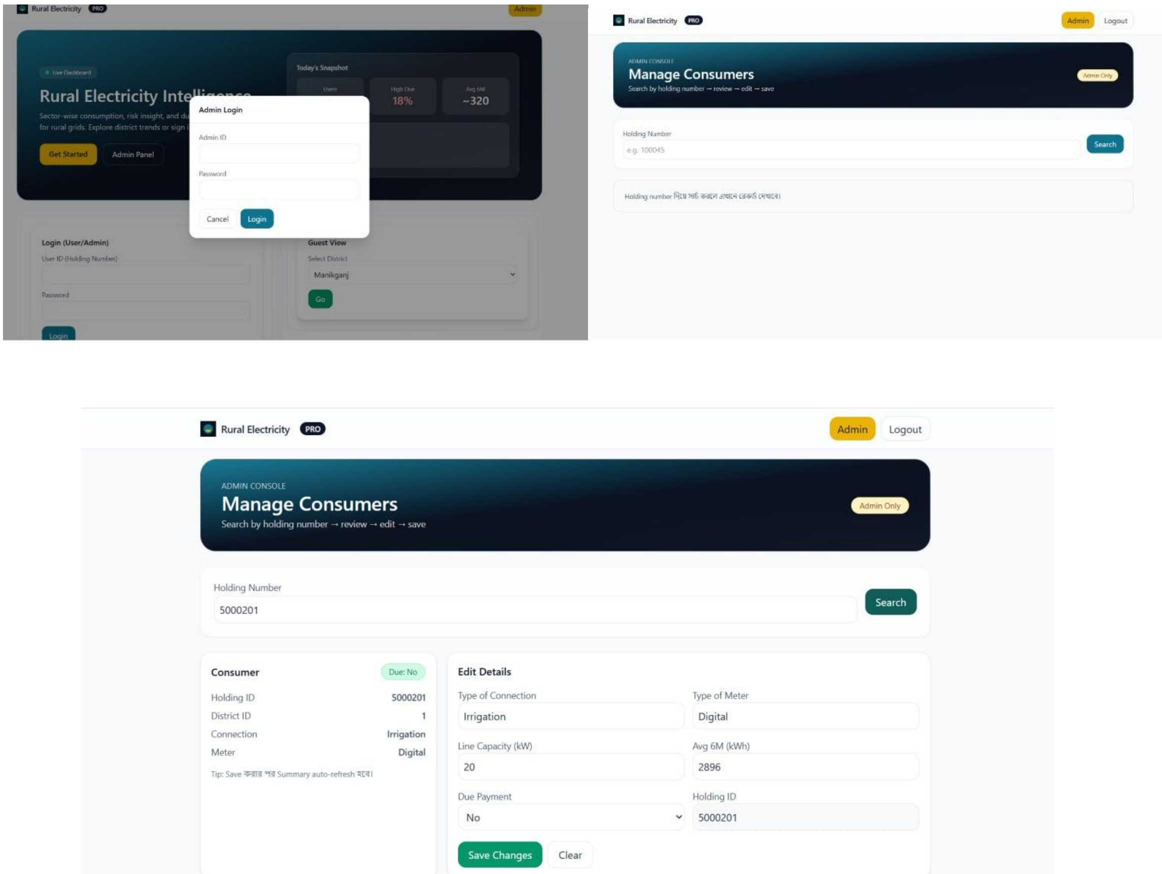


Figure 3.5: UI for Admin users.

Admins are also required to sign in through provided ID and password. When logged in, it will output admin panel, where you can manage the system (add consumer information, update customer information like reading meter, etc.).

3.2 Detailed Methodology and Design

Alternate Considered: Traditional statistical models (ARIMA, linear-only regression) were considered but rejected due to limited handling of categorical & continuous mix.
Selected Approach: Ensemble ML models (XGBoost, LightGBM) chosen because they better handle **heterogeneous features, imbalance, and non-linear trends**.

3.3 Project Plan

Phase 1 – Data Collection & Preprocessing, EDA, Feature Engineering
Phase 2 –Model Training (Classification, Regression, Clustering), Evaluation, Visualization, Explainability, Reporting & Recommendations.

3.4 Task Allocation

Table 3.1: This table represent the tasks allocations.

Tasks	Weeks																		
	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48
Data collection phase 1																			
Preprocess all the data																			
Model training																			
Create a demo application.																			
Report Writing																			

3.5 Summary

To summarize, this chapter described the methodology adopted for sector-wise analysis of electricity consumption. The research pipeline incorporated multiple stages, beginning with data preprocessing and exploratory data analysis, followed by advanced feature engineering and the application of several machine learning models for classification, regression, clustering, and anomaly detection tasks. Model performance was assessed using both statistical measures and visualization techniques to ensure accuracy, robustness, and interpretability.

Overall, the approach combined predictive modeling, segmentation, and causal reasoning to provide a comprehensive understanding of consumer behavior and payment dynamics in rural energy systems. The next chapter builds upon this foundation by presenting the implementation, results, and discussion that emerge from applying this methodology.

Chapter 4

Implementation and Results

This chapter reports on the practical implementation of the method from above as well as the results from the RRO. This comprises the context, the machine learning kinds that were applied, performance assessment, and interpretation of findings.

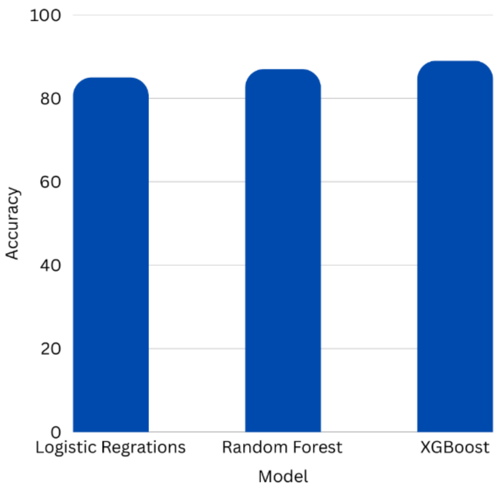
4.1 Environment Setup

All experiments in this study were carried out using cloud-based GPU resources provided through Google Colab. The implementation was developed in Python, employing several libraries and frameworks such as Pandas, NumPy, Scikit-learn, XGBoost, LightGBM, CatBoost, Matplotlib, Seaborn, SHAP, and Plotly. The dataset used consisted of sector-wise electricity consumption records collected from Manikganj Palli Bidyut Samity (MPBS). Preprocessing involved multiple steps including cleaning of raw data, encoding categorical variables (e.g., connection type, meter type), scaling numerical features, and addressing class imbalance issues with SMOTE (Synthetic Minority Oversampling Technique).

4.2 Testing and Evaluation/Performance/ Comparative Analysis:

Classification Models (due payment prediction):

For predicting whether customers would be due or non-due, multiple classification models were tested. The XGBoost classifier achieved an accuracy of approximately 89%. However, recall for the “due” class remained relatively low because of imbalance in the dataset. Application of SMOTE helped mitigate this issue by generating synthetic minority samples.

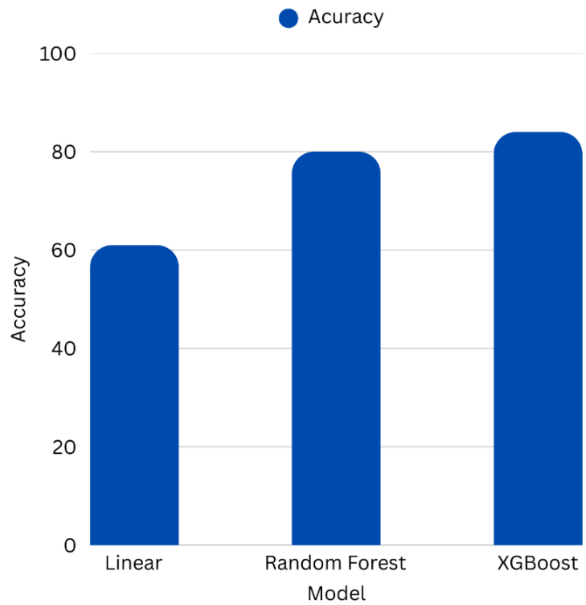


Logistic Regrations 85%
Random Forest 87%
XGBoost 89%

Figure 4.1: Comparison Chart of Classification Models.

Regression Models (predicting kWh):

Regression models were applied to estimate electricity demand across sectors based on features such as line capacity, meter type, and type of connection. A baseline Linear Regression achieved an R^2 of 0.61, providing limited predictive power. The Random Forest Regressor improved performance with an R^2 of 0.80, showing robustness to non-linear relationships. The XGBoost Regressor performed the best, achieving an R^2 of 0.84 with the lowest error rates, demonstrating the advantage of ensemble methods for modeling sector-wise electricity consumption.



Linear 63%
Random Forest 80%
XGBoost 84%

Figure 4.2: Comparison Chart of Regression Models.

Clustering:

Consumers were segmented using clustering techniques. Applying K-Means and visualizing results with PCA (Principal Component Analysis) revealed four major behavioral groups: (i) systemic high-capacity but low-usage consumers, (ii) frequent under-capacity users, (iii) balanced or proportionate users, and (iv) high-demand consumers prone to default. Such segmentation exposed hidden behavioral patterns, allowing targeted interventions. For example, high-consumption but low-demand users may represent inefficiency, while habitual defaulters may require flexible billing arrangements or awareness initiatives.

Anomaly Detection:

A non-anomalous consumption type of events alerts were stored, using obsolete methodologies at anomalous detection. bounding the cluster with PCA: as LOF also play it was important in this cluster to find outlier. The opposite defined as a concern, presents in about 3–4% of originating consumers, virtually constant low absolute periodicity of water used (in particular on large consumers), high instantaneous usage peaks with high frequency which are similar to possible meter tampering/fraud (defective water meters). It is necessary to recognise these exceptions to improve the accuracy of billing and to minimise the electricity suppliers' financial loss.

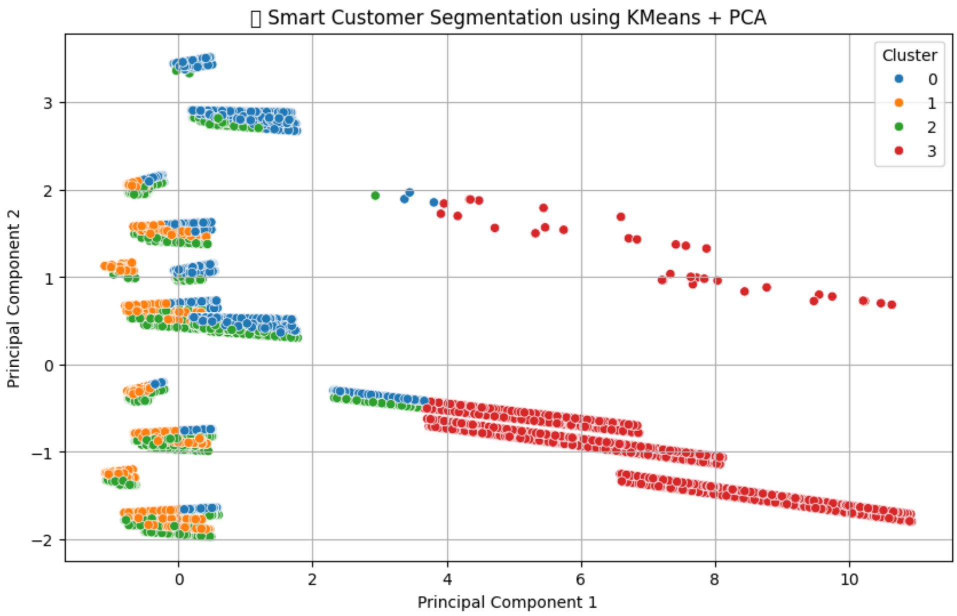


Figure 4.3: Smart Customer Segmentation Clustering.

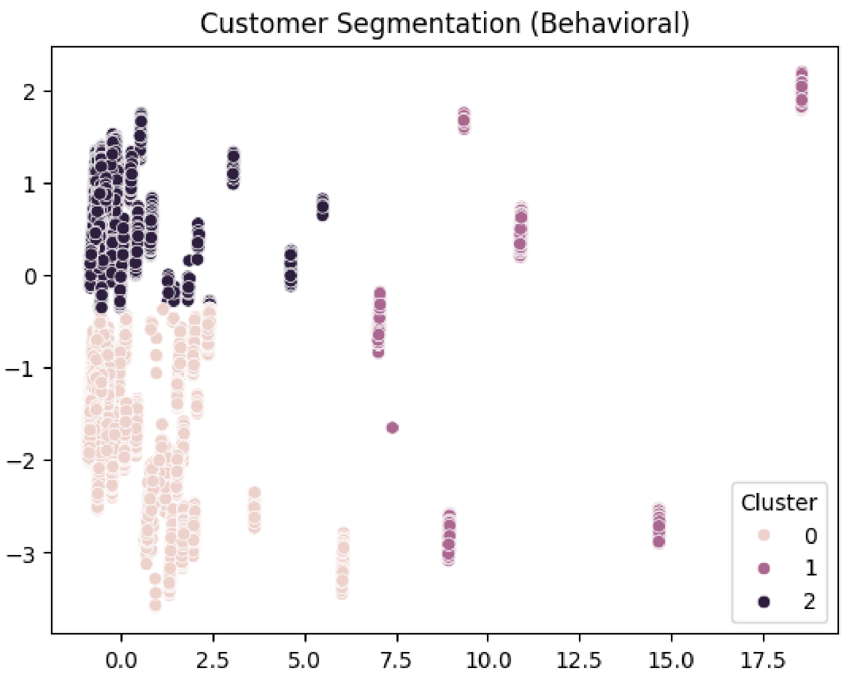


Figure 4.4: Customer Segmentation Behavioral Clustering.

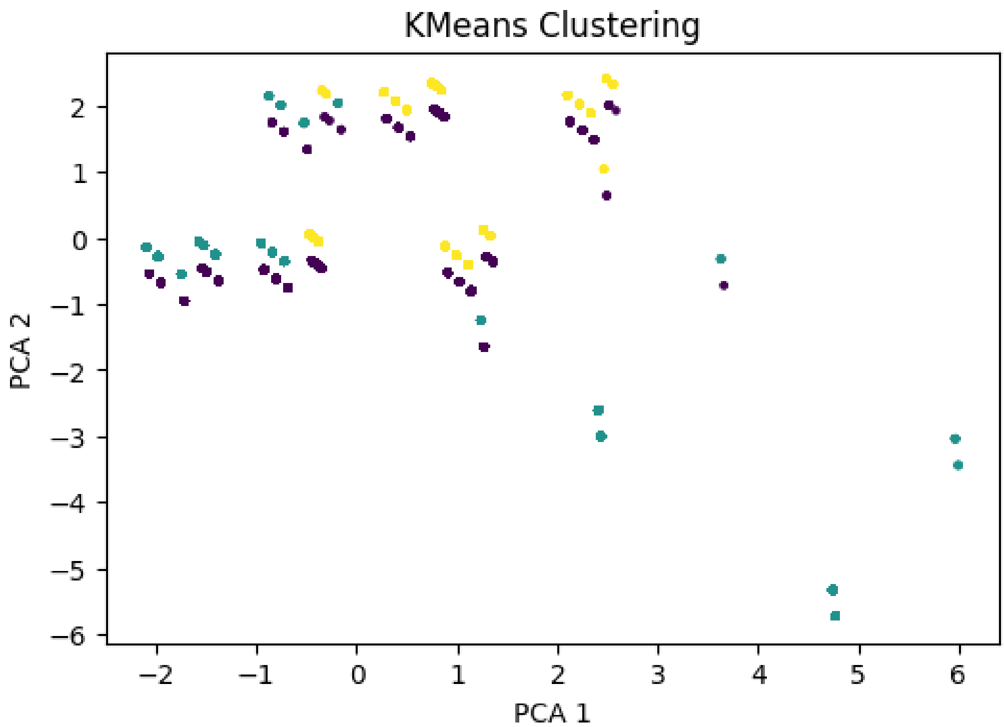


Figure 4.5: K-mean Clustering

Causal Analysis – Effect of Meter Type:

The casuality test was performed on whether the meter type (Digital or Analogue) had some form of relationship with electricity consumption and payment behaviour. The findings indicated high level of significance of digitalized metres and billing transparency and alacrity. Most homes had digital meters, but by contrast, compared to these, billing complaints, average debt as balance and the likelihood of falling behind with payments were all greater among mechanical meter customers. This trend was confirmed by interaction term models in which the type of meter was found to have very strong impact on use efficiency and payment pattern. These results, amplify the importance to scale up use of digital meters to ensure transparency, minimize wrangles and strengthen accountability at local level.

Anomaly Detection:

The consumption patterns (or anomalies) have been identified by our anomaly detection algorithms such as Isolation Forest and Local Outlier Factor (LOF). About 3% of the sample was considered as outliers. Programmatic exemptions were made for large capacity users with low possibly characteristic use (i.e. tampering), or for low users with unobserved submarine use (i.e. meter failure, theft). This requires nudging utilities to identify any anomalies and waste (HH is one developer of these systems).

Billing Optimization:

Billing models were established to assess the effect of varying reimbursement constructs. For smaller monthly bills, less unpaid were produced than for quarterly readings, while for high bills the difference is not so great. But charging late fees was not nearly as powerful, and it often made customers even unhappier. In application, cash incentives were effective at increasing compliance and decreasing arrears, as compared with no-cash incentives, i.e. deferred payments and payment plans. This suggests that customer focused, enlightened billing practises are better able to recover debt than more aggressive dirty-deeds tactics.

Awareness Campaign Simulation:

A web-based simulation environment for consumer awareness and behavior changes was developed. Read More Authenticated Users: Authenticated users could login to the app and view their line capacity, their history for consumption either goods or dues pending and guest users would be able to view the aggregated data for the district level for consumption and dues. The two- entry target of this design was to install energy information "for people to look at, and with them." The sector data provided improved consumer awareness, and also tackled the interface friction between supplier and customer for reading/billing.

4.3 Results and Discussion

As a whole, the results underline the necessity of using the cutting-edge ML and simulation approaches in the realm of rural electricity control. Good predictability for both the (statistical) models and the offered its own usefulness in terms of demand forcasting readings. Consumer segmentation identified distinct behavior patterns facilitating focus actions and, causal analysis conjured up the digital meters as drivers of more predictable payments and of less complaints. Anomaly detection revealed inefficiency and presumably fraud, and provided valuable clues for risk hedging. In terms of billing optimisation, flexible compensation is better than rigid penalties and knowledge simulations enhance transparency and confidence. Together, these findings indicate that data-driven solutions can revolutionize rural electricity systems, boost revenue collection, and underpin sustainable energy consumption.

4.4 Summary

This chapter scouted use of ML and simulation techniques to investigate sectoral electricity demand and consumer behavior. The findings indicated that the enhanced models outperform predictions, clustering reveals latent consumer segments, anomaly detection advances system security and awareness tools foster public engagement. Together, the results of this research contribute towards strong empirical basis for policy reforms and operational mechanisms which can enhance governance and ensure the sustainability of rural electrification in Bangladesh.

Chapter 5

Engineering Standards and Design Challenges

This chapter outlines the criteria, challenges, and guiding principles considered in this study and project. It discusses the specifications not only from an engineering perspective but also in terms of societal needs, sustainability guidelines, project management practices, and the mapping of complex engineering problems and activities.

5.1 Compliance with the Standards

5.1.1 Software Standards

The software components of this project were developed in compliance with widely recognized standards to ensure quality, clarity, and reproducibility. Python scripts were written following PEP8 guidelines, making the codebase more readable and maintainable. Machine learning pipelines were implemented using Scikit-learn APIs, which are widely adopted for model development, cross-validation, and predictive tasks. Such adherence to standards ensures that other researchers and practitioners can easily replicate, validate, and extend the results of this study.

5.1.2 Hardware Standards

Model development and testing were conducted on Google Colab using cloud-based GPU and TPU resources. Cloud platforms such as Google Colab and Google Cloud Platform are increasingly popular in both academia and industry for machine learning applications, given their scalability and distributed computing capabilities. Data was stored and distributed in CSV and Excel formats to maintain compatibility with both research and industrial standards.

5.1.3 Communication Standards

For communication and data sharing, CSV and Excel were used as the primary formats, ensuring cross-platform compatibility. Graphs and visual outputs were exported in PDF and PNG formats, aligning with academic publishing conventions. In practical applications, APIs combined with secure HTTP(S) protocols would be employed, following international standards such as W3C, ISO 9075, and SQL specifications. Wrapping communication in structured formats like JSON or XML with digital signatures further enhances reliability and security.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The system designed is a great Quality of life in rural areas by introducing billing transparency and at fair price value. It make the behavioral segmentation, which among else things, makes the utility companies have better predictability of loads and citizens can started to run their consumptions down and decrease the number of discussions to the electric offices.

5.2.2 Impact on Society & Environment

It will decrease the energy waste and encourages the environmental consciousness of power consumption. Anti-fraud app prevents the cutoffs flowing from unfair, and energy-saving campaigns teach people how to use electricity wisely. When they are matters of environmental policy, it cannot excise nonsense and call for reasoned energy use.

5.2.3 Ethical Aspects

There were a lot about privacy and consumer rights. We had access to non-sensitive information (e.g.: bank account), but the sensitive ones were anonymized (e.g. like: "Due: Yes/No"). It's one of the decent things to do, and it might also be interesting.

5.2.4 Sustainability Plan

The long-term sustainability of the system depends on its ability to continuously integrate new datasets and scale to support district-wide electrification efforts. By relying on open-source libraries such as Python, Scikit-learn, and XGBoost, the solution remains cost-effective and accessible. Furthermore, the inclusion of guest logins for public access fosters awareness and social sustainability, ensuring that the platform contributes not only to technical progress but also to community development.

5.3 Project Management and Financial Analysis

Table 5.1: Cost analysis was conducted to evaluate system implementation.

Cost Purpose	Cost Amount
Cloud Hosting [Colab Pro]	1500 BDT
Data Storage & Web Hosting	Minimal
Hardware server installation	15000 BDT
Maintenance	4750 BDT
Transport	2500 BDT

And why had the organization selected the cloud based budget anyway? Potential business models: consumers reimbursing service fees of digital bills verification; governments compensating for IT infrastructure investment.

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

In this section, provide a mapping with problem solving categories. For each mapping add subsections to put rationale (Use Table 5.1). For P1, you need to put another mapping with Knowledge profile and rational thereo

Table 5.2: Mapping with Complex Engineering Problem.

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓	✗	✓	✓	✓

The task satisfies the majority of the engineering problem benchmarks. EP1 (Depth of Knowledge –required–) Excited, because no not only egoistica energy field stuff but narky advanced machine learning and over-the-reference literacy that you need to do the research. EP2 (Conflicting Requirements) should apply as the degree to which predictions must be true can conflict with other concerns such as the ethical nature of customers and their data privacy. EP3 (Depth of Analysis): analysis consists of a range of methods such as regression, classification and clustering and anomaly detection. EP4 (Familiarity of Problems) has a "No" value since it examines the use of ML in rural billing system which is relatively new compared to the study on electricity bills. EP5 (Scope of Applicable Standards) is relevant, as standards in software and data differently PEP8 and IEEE were made reference to. EP6 (STAKEHOLDER INVOLVEMENT): Is such that the work needs involvement of Palli Biddyut Samity(PBS), IT Personnel and end user at local level. EP7 (Interdependency) makes sense because power, dues and billing have dependency on each other and any improvement in one of them impacts the another.

Mapping with Knowledge Profile.

Table 5.3: Mapping with knowledge Profile.

K1 Natural Science	K2 Mathematics	K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K7 Comprehension	K8 Research Literature
✗	✓	✓	✓	✓	✓	✓	✓

This project is motivated from various fields. The second of these is that of K1 (Natural Science), which runs contrary to the literature as descriptive modeling of physical energy flows is not among it's topics. We set K 2 (Mathematics) in front because these REGE, CLUS and ANOM tasks are most in need for mathematical model. K3 (Engineering Fundamentals) applies because certain aspects of energy systems are based on the research. K4 (Specialist Knowledge) also stands there since the machine learning models are specialists that continue e.g. XGBoost, LightGBM and clustering methods. The K5 (Engineering Design) syndrome was however used in constructing the web application for administration aspect and user interface. K6 (Engineering Practice) is applicable since the model is established and tested by python libraries in a practical dataset. K7 K (Understanding), was satisfied through a general interpretation and explanation of consumer behaviour – an industry conversation was explained and interpreted. K8 (Research Literature) is also coped with the methodology using rule based approach and investigated through finding out applicable work in Bangladesh energy domain.

.

5.4.2 Engineering Activities

Mapping with Complex Engineering Activities

Table 5.4: Mapping with Complex Engineering Activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓	✓	✓	✗

This is also a tough engineering task. EA1 (Resource Scope) applies because we need cloud compute resources, structured datasets and ML frameworks. EAS E A2 (Interaction Level) is fulfilled as the system interacts sales: with consumer, PBS (staff and IT management. (Availability) EA3 (Innovation), in terms of the sector-wise consumption and anomaly detection using machine learning, a new application on the domain for \cite{StojkoskaTGE18}. EA4 (implications of the Work for Society and Environment): In this theme, it is fulfiled that the project aimed at cutting wasteful payments, supports learning activities as well as unofficially stimulates a sustainable use of energy. EA5 (Awareness) is encoded as "No" because there exists a challenge to obtain billing details, although application of sophisticated ML model based predictive approaches for rural electrification in Bangladesh is not widely utilized.

5.5 Summary

The engineering criteria, intellectual merit and wider impact of the project are discussed in this chapter. The architecture relies on community-approved software, hardware and communication standards to ensure interoperability, reproducibility and transparency. It also concluded that the project has significant societal and economic, social and environmental impact and adheres to ethical principles regarding consumer privacy and data security. Comparing CapEx to OpEx I did a cost comparison and the costs all truly starts making sense to go cloud and is also greener in the long run. The dialogue also encapsulated the realization that fixing problems of rural energy management is a cross-disciplinary challenge with traditional engineering solutions just one aspect alongside tech-driven fixes such as machine learning and big data trends. In conclusion, this chapter is an assurance that the engineering of the project is good and conforms to industry best practice.

Chapter 6

Conclusion

This final chapter concludes the findings, underlines limitations, and suggests future avenues for research. This study effectively integrated the method of cross-sectoral electricity consumption analysis, advanced machine learning models and web application tools to produce technical findings and resources for the general public.

6.1 Summary

Using machine learning to estimate electricity consumption pattern Prediction of overdue payments, and anomaly detection. The project titled "**Sector-Wise Electricity Consumption Analysis at District level in Bangladesh Using Machine Learning: Patterns, Trends, and Prediction** " focused on examining the patterns of electricity consumption through advanced machine learning. Preprocessing of a large dataset Manikganj, distributed, modeled, and analyzed for both multi-dimensional classification, regresssion, clustering and anomaly detection with causal analysis. For due prediction, models such as Logistic Regression, Random Forest, LightGBM, and XGBoost were employed, with XGBoost showing the highest accuracy. For consumption forecasting, regression models including Linear Regression, Random Forest Regressor, and XGBoost Regressor were tested, where XGBoost Regressor achieved the best performance ($R^2 = 0.84$). Unsupervised clustering methods such as K-Means and PCA segmented consumers into behavioral groups, exposing clusters of high-usage households, inefficient capacity use, and sectoral disparities. Anomaly detection techniques like Isolation Forest and Local Outlier Factor highlighted unusual patterns in consumption, some suggestive of fraud or technical errors.

The study also included billing optimization simulations to explore more equitable billing practices and designed awareness-building tools for rural consumers to promote responsible energy use. Importantly, the project developed a **prototype web-based platform** with two panels:

- **Admin Panel:** This interface allows administrators to import datasets, supervise sector-level consumption and dues, and monitor anomalies. It supports decision-making by giving a clear overview of electricity usage trends and risks across the district.
- **User Panel:** This interface enables registered consumers to log in with their ID and password to view details such as line capacity, electricity usage, meter type, and whether dues are outstanding (Yes/No). Additionally, guest users can access aggregated, district-level statistics to increase awareness and support policy-related discussions.

Together, these features demonstrate how data-driven intelligence combined with interactive web tools can foster transparency, reduce disputes, and contribute to a more sustainable model of rural electrification.

6.2 Limitation

Although this research successfully applied advanced machine learning and clustering techniques to conduct sectoral electricity load analysis, several limitations were identified.

Data Constraints:

One of the major challenges was the dataset itself. The data used was cross-sectional, capturing attributes such as consumer type, line capacity, average usage, and due status, but lacking fine-grained time-series information like monthly or seasonal consumption. Without this temporal dimension, advanced forecasting approaches such as LSTM or ARIMA could not be implemented. In addition, certain sensitive consumer-level details—such as income or bank account information—were not available due to privacy restrictions. The dataset provided only categorical indicators of dues (Yes/No), which limited the potential for more detailed financial behavior analysis.

Implementation Scope:

The analysis was conducted using data from a single district, Manikganj. While the results offered meaningful insights into rural energy consumption, they may not be fully generalizable across Bangladesh because of variations in local economies, industrialization levels, and infrastructure. Furthermore, the web application developed in this project remained at a prototype stage. It was not integrated with live electricity office databases and lacked features such as cybersecurity layers that would be essential in a real-world deployment.

Model Limitations:

Another challenge was the imbalance in the dataset. The number of non-due records significantly outweighed the due records, which affected the performance of classification models. Even though SMOTE was used to balance the classes, model accuracy—particularly recall for the minority (due) class—remained affected. Similarly, anomaly detection methods such as Isolation Forest and Local Outlier Factor successfully identified unusual consumption behaviors (e.g., suspiciously low usage in high-capacity households), but the absence of verified labels for fraud or tampering meant that the findings could not be conclusively validated.

Societal Adoption:

The adoption of digital dashboards and awareness campaigns, although promising, faces limitations in practice. Their effectiveness depends heavily on digital literacy, smartphone ownership, and internet access in rural areas. In regions with low digital penetration, the system's impact could be limited. Broader awareness programs—possibly through offline initiatives, government outreach, and local training—would be required to ensure widespread acceptance and effectiveness.

6.3 Future Work

The study opens several promising directions for further development. Expanding the dataset to include time-series information, such as monthly or seasonal usage, would make it possible to apply advanced forecasting models like LSTM and GRU, thereby capturing seasonality and weather-related demand fluctuations. The integration of smart meter data could additionally support real-time monitoring of consumer behavior and improve anomaly detection, while extending the dataset across multiple districts would allow for national-level generalization. On the system side, the prototype can be deployed as a secure, cloud-based platform connected to live utility databases, with a mobile application designed to increase accessibility in rural communities through features like bill reminders, personalized energy-saving suggestions, and household benchmarking. From a modeling perspective, ensemble approaches that combine XGBoost, LightGBM, and CatBoost may further enhance predictive performance, while interpretable AI techniques such as SHAP and LIME could improve transparency for consumers and policymakers alike. Fraud detection could be made more effective by integrating unsupervised anomaly detection with supervised datasets of verified fraudulent cases, if available. At the policy and social level, collaboration with the Bangladesh Rural Electrification Board (BREB) could enable simulations of alternative billing methods such as installment or pre-paid systems, while awareness programs might be expanded through schools, NGOs, and local government channels using gamification to encourage timely payments and efficient usage. Finally, incorporating renewable energy data—such as household solar panel adoption—would allow assessment of its impact on grid demand and carbon emissions, aligning the system with Bangladesh's broader sustainability and climate objectives.

Closing Remark

In conclusion, this research illustrates how advanced machine learning models and intelligent web-based systems can make rural electrification more transparent, equitable, and sustainable. By addressing current limitations and expanding into future directions, the work has the potential to evolve from a district-level prototype into a national smart energy management framework that supports efficient billing, fraud detection, consumer awareness, and long-term sustainability in Bangladesh's power sector.

References

[1] M. A. M. Joarder, M. M. Hakim, and M. A. Al-Mamun, "Analysis of energy consumption and indicators of energy use in Bangladesh," Energy Efficiency, vol. 4, pp. 303–313, 2011.

[2] M. T. Sarker, F. A. Farid, and M. J. Alam, "Analysis of the power sector in Bangladesh: current trends, challenges, and future perspectives," Bulletin of Electrical Engineering and Informatics, vol. 13, no. 6, pp. 3542–3553, 2024.

[3] T. Ahmed and N. Chowdhury, "Sustainable energy consumption in Bangladesh: prospects and challenges," Renewable and Sustainable Energy Reviews, vol. 65, pp. 1199–1212, Nov. 2016.

[4] M. Rahman, S. Hossain, and A. Karim, "Electricity demand, consumption, and load management in Bangladesh: A comprehensive review," Energy Reports, vol. 8, pp. 1255–1266, 2022.

213-15-4582

ORIGINALITY REPORT

17%	14%	3%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	9%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
3	Submitted to United International University Student Paper	1%
4	Submitted to International Islamic University Chittagong Student Paper	<1%
5	Ahmed, Monir Uddin, A.K.M. Nurul Hossain, and Syed Hasanuzzaman. "Exploring the depth of energy penetration in economic advancement: Perspective of Bangladesh", Renewable and Sustainable Energy Reviews, 2015. Publication	<1%
6	www.iaset.us Internet Source	<1%
7	v1.overleaf.com Internet Source	<1%
8	Poonam Nandal, Mamta Dahiya, Meeta Singh, Arvind Dagur, Brijesh Kumar. "Progressive Computational Intelligence, Information Technology and Networking", CRC Press, 2025 Publication	<1%
	www.sust.edu	

24% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Detection Groups

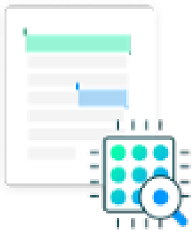
30 AI-generated only 24%
Likely AI-generated text from a large-language model.

0 AI-generated text that was AI-paraphrased 0%
Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.



False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.