Integration of Machine Learning Algorithm and Geospatial Mapping Technique for the Electric Distribution Power Outage Monitoring

BY

NWAEZE Daniel Chinedu

XXXXXXXXX

SUBMITTED TO:

THE DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

UNIVERSITY OF LAGOS, AKOKA

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE (B.Sc. Hons) IN ELECTRICAL & ELECTRONICS ENGINEERING

SUPERVISED BY

Dr. Peter O. Oluseyi

August 2024

CERTIFICATION

This is to certify that this project was carried out by Nwaeze Daniel Chinedu with the matriculation number XXX in the Department of Electrical and Electronics Engineering, Faculty of Engineering, University of Lagos under the supervision of Dr. Peter O. Oluseyi.

Nwaeze Daniel Chinedu
Matriculation Number:

Date

Author

Dr. Peter O. Oluseyi

Date

Supervisor

Dr. (Mrs.) Gbenga-Ilori

Date

Project Coordinator

Prof. Akinbulire (H.O.D)

Date

DEDICATION

To God, my parents, siblings, mentors and colleagues. Your guidance, support, and passion inspired this work.

ACKNOWLEDGEMENT

First and foremost, I express my deepest gratitude to the Almighty God for granting me the strength, wisdom, and perseverance to undertake and complete this project.

I would like to extend my sincere appreciation to my supervisor, Dr. Peter O. Oluseyi, whose guidance, support, and encouragement have been invaluable throughout this endeavour. His expertise and mentorship have greatly enriched this research.

I am also thankful to Dr. (Mrs.) Gbenga-Ilori, the project coordinator, as well as Engr. Adenola for their assistance and oversight, which have been crucial in ensuring the successful execution of this project.

Furthermore, I acknowledge Prof. Akinbulire, the Head of the Department of Electrical and Electronics Engineering and the entire staff in the department, for providing the necessary resources and facilities essential for the completion of this work.

My gratitude also goes to the management and staff of Ikeja Electric Distribution Company (IKEDC) for their cooperation and assistance in providing access to crucial data and information. I am grateful to my colleagues and friends for their support and encouragement throughout this academic journey.

Lastly, I express my heartfelt appreciation to my family for their unwavering love, understanding, and encouragement, which have been my source of strength throughout this endeavour.

Nwaeze Daniel Chinedu

ABSTRACT

The power outage events are global phenomenon. It often impedes sustainable development and affects the productivity of various sectors of the economy. To reduce the unscheduled outage in the distribution network; this project was developed to enhance outage monitoring, prediction and management by integrating state-of-the-art technology. There are very few works in this area of research. Thus, the previous research gaps included the lack of proactive outage prediction and underutilisation of Geographic Information Systems (GIS) in electrical distribution topology.

The methodology involved developing predictive models using historical outage data, weather details, and system operations information. This is achieved by visualising GIS maps with OpenStreetMap, guided by interdisciplinary principles from electrical engineering, data science and spatial analysis. This study was validated with data from Ikeja Electric Distribution network in Lagos. The choice of this case study is due to its extensive network and diverse infrastructure, providing rich data for predicting power outages and visualizing them using GIS. This approach helps in developing effective models to enhance the resilience and reliability of the distribution network. The results obtained included 74% predictive accuracy in forecasting outages 24 hours in advance, 85% accuracy in classifying outages (when calibrated with actual historical fault data). Thus, with the web integration with the GIS mapping, a significant enhancement of the outage monitoring capabilities was recorded; this suggests an improvement in the distribution network's resilience. The project underscored the importance of proactive prediction and recommended widespread adoption of GIS-based solutions. It is recommended that the further research (on this study) should focus on refining predictive models and expanding GIS-based solutions to extend the software application to other distribution network topologies in Nigeria.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF EQUATIONS

LIST OF FIGURES

LIST OF ABBREVIATIONS

- ADN - Active Distribution Network
- ANN - Artificial Neural Network
- AUC - Area Under Curve
- CNN - Convolutional Neural Networks
- CSS - Cascading Style Sheets
- DER - Distributed Energy Resources
- DMS - Distribution Management Systems
- EDA - Exploratory Data Analysis
- GIS - Geographic Information Systems
- GLM - Generalised Linear Model
- HTML - HyperText Markup Language
- IDE - Integrated Development Environment
- IKEDC - Ikeja Electric Distribution Company
- JS - JavaScript
- ML - Machine Learning
- NESO - Nigeria Electricity System Operator
- NLP - Natural Language Processing
- NLSI - New Line Stability Index
- NumPy - Numerical Python
- OMS - Outage Management Systems
- OSM - OpenStreetMap
- RNN - Recurrent Neural Networks
- SDLC - Software Development Life Cycle
- SGSTF - Structure, Granularity, Scope, Temporality and Faithfulness

- SO - System Operator
- SVC - Support Vector Machine
- TCN - Transmission Company of Nigeria
- TF-IDF - Term Frequency Inverse Document Frequency
- TS - TypeScript

CHAPTER ONE

INTRODUCTION

Background of the Study

Power distribution networks are critical infrastructures that ensure the delivery of electricity from transmission plants to end consumers. These networks are the backbone of modern society, facilitating the operation of industries, businesses, healthcare facilities, and households. The resilience of these systems is paramount, as any disruption in power supply can lead to significant economic losses, compromised safety, and diminished quality of life.

Power outages, whether planned or unplanned, are a global challenge with wide-ranging impacts. They can result from various factors, including weather conditions, technical faults, equipment failures, and human errors. The frequency and severity of these outages underscore the importance of robust outage management and mitigation strategies. In the context of sustainable development, reliable power distribution is essential for economic stability, public safety, and the efficient functioning of essential services.

The field of power distribution has seen substantial advancements in recent years, driven by the integration of modern technologies such as Geographic Information Systems (GIS), data analytics, and machine learning. These technologies have opened up new possibilities for enhancing the resilience and efficiency of power networks. GIS, in particular, plays a crucial role in visualising and analysing the spatial aspects of power distribution, enabling more effective outage management and infrastructure planning.

Despite these advancements, challenges remain, particularly in regions like Nigeria, where power distribution networks are often strained by growing populations and economic activities. The focus of this research is on addressing these challenges by leveraging predictive modelling and GIS to enhance the monitoring, prediction, and management of power outages. Specifically, this project aims to bridge the gap in proactive outage prediction and the underutilization of GIS in the Nigerian context, with a case study focusing on the Ikeja Electric Distribution network in Lagos. By integrating these technologies, the project seeks to develop innovative solutions that will improve the resilience and reliability of power distribution in urban environments.

Problem Statement

Power disruptions pose a widespread challenge for electricity distribution systems globally, stemming from diverse causes such as climate-related factors, technical malfunctions, and human-related issues. These outage events can lead to significant financial repercussions [1], including food spoilage [2] or critical health emergencies in medical facilities, especially when backup power sources fail [3].

A 2016 economic evaluation, which shows a startling monthly expenditure of ₦45,811,859 for power generation within the relevant industry [4], clearly illustrates the severity of these interruptions. The financial strain placed on firms not only compromises their capacity to remain profitable, but it also affects consumers directly by driving up production costs and product prices.

With Lagos's ever-growing population and dynamic business environment, the instability of the electricity supply has emerged as a major barrier to sustainable development in the city. Beyond the financial consequences, power outages jeopardise public safety, interfere with essential services in medical facilities, and lower citizens' quality of life in general [5].

Even though power outages are acknowledged as a widespread problem, most research conducted in Nigeria so far has concentrated on post-event analysis, classifying failures based on frequency and addressing the type of disruptions experienced rather than proactively predicting them [6]. The efficacy of outage management techniques is restricted by this retroactive approach, which also impedes the creation of predictive metrics that are essential for enhancing the resilience of the power system as a whole.

Furthermore, there is a clear lack of integration in Nigeria's mapping and visualisation of outage areas using Geographic Information Systems (GIS) [7]. Although GIS has shown to be extremely helpful in outage management throughout the world, its underutilisation in Nigeria leads to a disjointed understanding of the causes of outages, which impedes the development of comprehensive mitigation and preventive solutions [8].

The crucial need for an innovative and comprehensive strategy to power outage management in Lagos, Nigeria is emphasised by the lack of preventative measures, the restricted use of GIS, and the economic ramifications of frequent power interruptions.

These issues underscore the urgent need for research and solutions aimed at enhancing outage prediction, safety monitoring, and stakeholder communication. By addressing these challenges, efforts can be directed towards improving the resilience and reliability of the electrical distribution system in Lagos, Nigeria.

In line with the United Nations Sustainable Development Goal (SDG) 7, aiming for universal access to affordable, reliable, sustainable, and modern energy for all by 2030, it acknowledges the significance of reliability as a crucial aspect within the access definition (SDG 7.1).

Aim and Objectives of the Study

The project aims to integrate geospatial mapping and predictive modelling to enhance power outage monitoring by accurately predicting outages before they occur, understanding outage patterns and locations, which aids in initiating pre-emptive safety measures.

To achieve these aims, the following objectives shall be pursued:

I. Design and develop predictive models to accurately forecast both planned and unplanned high impact, low probability (HIPL) outages, leveraging network fault logs.

II. Create spatio-temporal geospatial maps of outages using OpenStreetMap, providing stakeholders with a comprehensive understanding of the geographical impact of electricity disruptions.

III. Implement an outage monitoring web system that integrates predictive models and geospatial maps to provide outage prediction, classification and monitoring, enhancing operational efficiency in managing power disruptions.

Significance of the Study

The significance of this study lies in its potential to revolutionise power outage management by integrating geospatial mapping with predictive modelling, providing a proactive framework for anticipating and mitigating outages. This approach enhances predictive accuracy, with models that forecast outages with 74% accuracy, thereby improving power supply reliability and minimising economic losses and safety risks. The use of Geographic Information Systems (GIS) offers a detailed and spatially accurate understanding of outage hotspots, enabling targeted interventions and infrastructure improvements, which is particularly valuable in areas like Lagos. By validating these models with data from the Ikeja Electric Distribution network, the study provides actionable insights that can be directly applied in real-world scenarios and serves as a blueprint for other distribution networks.

This research supports sustainable development by ensuring a more stable electricity supply, protecting vital services, reducing economic disruptions, and improving citizens' quality of life. Additionally, it lays the groundwork for future research, encouraging the continued integration of interdisciplinary approaches to address challenges in the power sector. The study's findings contribute to both academic knowledge and practical applications, making it a significant advancement in outage monitoring and management.

Scope and Limitations of the Project

The project encompasses the development of methodologies and tools for improved power outage management, leveraging data from the Ikeja Electric Distribution Company (IKEDC) in Lagos, Nigeria, spanning from October 2, 2021, to December 2, 2023. Additionally, Visual Crossing's weather data, Nigeria Electricity System Operator (NESO) data, and OpenStreetMap are incorporated, providing a broader scope of applicability beyond the specific distribution company's dataset. This approach ensures that the developed methodologies and tools can be generalised for implementation across other distribution companies.

While IKEDC serves as the primary dataset source, the methodologies developed aim to be transferable and applicable to diverse distribution companies across Nigeria. The project focuses on addressing weather-induced, technical-induced, and operational/human-induced outages, aiming to provide scalable solutions applicable to various distribution networks.

The predictive model developed aims to forecast the likelihood of an outage occurring within a day (24-hour period).

While striving to offer robust solutions for power outage management, the project confronts some acknowledged limitations. Accessing comprehensive historical outage data presents a significant challenge, potentially impacting the predictive model's accuracy and effectiveness. The dataset's coverage from 2021 to 2023, raises concerns about developing long-term predictive capabilities beyond that timeframe. Moreover, infrastructure-related challenges, encompassing data collection mechanisms and system integration, constrained the project's scope and implementation.

Furthermore, the project's aspiration for generalisability prompts recognition that insights derived from IKEDC data might not perfectly mirror other distribution companies' circumstances. Additionally, the reliability and accuracy of weather data obtained from Visual Crossing may vary, potentially compromising the model's predictive capabilities.

Human factors further compound the complexity, as efforts to address operational and human-induced outages must contend with the inherent variability of human behaviour and operational practices. Despite these acknowledged limitations, the project remains committed to providing valuable insights and tools for enhancing power outage management. Recognising the necessity for ongoing refinement and adaptation based on evolving data and operational contexts underscores the project's commitment to resilience and reliability in electrical distribution systems.

Report Outline

This report is structured into seven sections. The introduction presents an overview of the project's objectives. Section 2 examines relevant literature and its alignment with the project's aims. Section 3 details the design of the system and the methodology for achieving project goals. Section 4 presents the results and discusses their implications. Section 5 summarises the findings, offers conclusions, and makes recommendations for future work. Additionally, section 6 includes the list of references consulted during the project and section 7 contains the appendix, providing supplementary information.

In summary, this project integrates advanced techniques like geospatial mapping and machine learning to significantly enhance the accuracy of power outage predictions and implement proactive monitoring. The research involves developing predictive models using historical outage data, weather patterns, and system operation information, alongside Geographic Information Systems (GIS) for visualising outage locations. These innovative approaches aim to revolutionise power outage monitoring, improving the resilience of the electrical distribution system in Nigeria, particularly focusing on Lagos.

CHAPTER TWO

LITERATURE REVIEW AND THEORETICAL FRAMEWORK

Overview of Existing Relevant Literature/Research

Power outage management and prediction have garnered significant attention in research communities worldwide, reflecting the critical importance of reliable electricity supply for societal and economic activities [9-11]. In the Nigerian context, where power disruptions are pervasive and pose substantial challenges, recent studies have explored the impact of electricity outages on the socio-economic state of the country.

Research conducted to ascertain the relationship between electricity supply and a country's economic growth and development reveal a unidirectional relationship without a feedback effect between labour and electricity supply [12].

In a study conducted, the correlation between electricity provision and economic advancement in Nigeria from 1970 to 2009 was explored [13]. The findings indicate that per capita Gross Domestic Product (GDP), past electricity provision, technology, and capital are the key factors influencing economic progress in Nigeria.

In their examination conducted in 2012, [14] investigated the correlation between electricity power and unemployment rates in Nigeria. They used the ordinary least square regression model to analyse the impact of electricity power outputs, supply, and consumption on addressing the high unemployment rates in Nigeria. The study period spans from 1970 to 2005. Their findings indicate that the supply of power to the industrial sector in Nigeria was lower compared to residential consumption. Furthermore, their research revealed that the primary cause of unemployment in Nigeria could be attributed to the insufficient and unstable power supply to the industrial sector.

The development of a robust strategy for mitigating outage risks [8] becomes crucial for utility managers, enabling them to devise protective and preventive measures aimed at minimising the occurrence of electrical disruptions. Utilising outage forecasting models and mapping systems emerges as a valuable tool in understanding both historical and future trends in outage events, thereby facilitating the formulation of effective preventive and mitigation strategies.

In the bustling city of Lagos, Nigeria, where having a steady power supply is crucial for daily life and business, dealing with frequent power outages is a significant challenge.

This literature review takes a closer look at recent research efforts aiming to improve how we manage these outages. By combining predictive modelling and geospatial visualisations, researchers are exploring innovative ways, like Machine Learning (ML) and Geographic Information Systems (GIS), to better understand and predict patterns in power interruptions, in order to anticipate them proactively.

While power disruptions across the continent are frequent, unpredictable, and pose both economic and social challenges, there's a limited understanding of their scope and patterns among the general population and in existing literature. As highlighted by [15], the absence of detailed data and frameworks on the causes and frequency of outages hinders in-depth analysis.

This project aims to fill this gap, focusing on Nigeria and utilising available data and frameworks. Current studies on power outages in Nigeria typically analyse causes and impacts retrospectively rather than proactively preventing them. For instance, [6] concentrates on post-event analysis of electricity outages in Lagos, using statistical methods to categorise outages by frequency. However, there's a notable absence of proactive measures, such as predicting where and when outages might occur in advance.

In a different approach, [16] explores the effects of power outages on micro-sized businesses to guide policymakers in fostering business development. Their study relies on survey data collected through purposive sampling, emphasising understanding business owners' perspectives on the extent and implications of outages rather than modelling outage patterns.

In [15], a Monte Carlo Analysis framework was employed to assess the environmental and economic impacts of backup diesel generators in Sub-Saharan Africa due to frequent power outages. Their findings indicated that relying on backup generators resulted in more adverse outcomes compared to a consistent electricity supply from the grid.

Similarly, [17] tackled voltage instability and collapse in electricity reliability, introducing the "new line stability index" (NLSI_1) for predicting voltage collapse on the Nigerian grid. This index considered switching logic based on voltage angle differences between two load buses.

In a parallel effort, [18] dedicated their research to predicting electrical power generation in Nigeria, crafting sophisticated regression and artificial neural network models.

While these contributions are valuable to the field, they generally lack direct emphasis on the nature of outages. Notably, none of these studies implemented machine learning algorithms to proactively predict faults leading to outages. Additionally, there's a notable absence of geospatial mapping of outage causes, which could contribute to a more comprehensive understanding of these disruptions.

Machine Learning Based Solutions

In tackling the primary challenge at hand, a comprehensive review extended beyond the confines of Nigeria, encompassing relevant literature from authors with a broader perspective. The exploration of machine learning methodologies takes centre stage in addressing outage prediction, responding to a spectrum of causes outlined below:

Climate and Weather Induced Outage

Weather extremes stand out as prominent catalysts for power outages globally, with [19] underscoring their significant impact. This study highlighted the prevalence of extreme weather as a major contributor to outages, with lightning alone causing 70% of disruptions in Malaysia. Employing logistic regression and decision tree models on weather data and outage reports, a superior Area Under Curve (AUC) result of 79% was achieved by the decision tree model, albeit with potential overfitting concerns due to a limited three-month dataset.

In the United States, [20] emphasised the staggering economic costs of weather-related power outages, amounting to tens of billions annually. Their approach involved developing a gradient boosting model, achieving an r-squared of 0.61 for outage prediction across various storm types, excluding thunderstorms. This success highlighted a robust correlation between weather data and outage forecasts, showcasing the model's efficacy in anticipating disruptions.

Addressing outages induced by snow and ice storms, [21] introduced models demonstrating the effectiveness of Generalised Linear Models (GLMs) like logistic regression for extreme weather events. Machine learning approaches, particularly adept at predicting lower impact events, were instrumental in characterising the spatial distribution of power outages. Other studies [22-24] further associated weather conditions such as typhoons and hurricane Sandy with electrical power system failures, employing robust random forest techniques to mitigate overfitting risks.

Technical Induced Outage

While prevailing research indicates a correlation between most outages and weather-related events, technical challenges persist as contributors to outage patterns.

In the pursuit of resolving recurring power outages, enhancing distribution network reliability, and mitigating customer complaints, [25] delved into concealed factors influencing frequent outages. Although weather events emerged as the primary concern, the age of equipment, particularly feeders leading to feeder faults, was identified as a significant factor. The study observed a positive correlation between equipment age and failure rate, noting a decline after a certain age. This phenomenon was attributed to the gradual exposure of equipment issues, leading to timely replacements. The longer the service life, the better the equipment quality, resulting in a lower probability of failure.

In a parallel effort, [26] pioneered a machine learning model addressing system operability disruption and equipment failure rate, contributing to a 5% power outage revenue loss in the United States from 2000 to 2016. The random forest model exhibited superior performance, achieving an accuracy of 79.8% and an r-squared of 0.76, surpassing the next-best Artificial Neural Network (ANN) model with an accuracy of 75.2% and an r-squared of 0.68. This demonstrated the efficacy of machine learning in comprehending and mitigating technical-induced outages.

Operational and Human Induced Outage

In the investigation of power outage incidents, [27] utilised Eaton's Blackout Tracker and similar reports to categorise operational-related outages into planned events or human errors. Planned outages encompassed scheduled work, public safety requests, or customer-initiated requests, while human errors comprised hacking, accidents, or theft/vandalism.

In a parallel initiative, [28] harnessed a machine learning random forest model to predict operational power system outages. Leveraging historical data on power system conditions, including voltage and current, the model accurately identified potential three-phase power failures with an impressive accuracy of 99.542%. Furthermore, it successfully determined the occurrence date of power failures based on given current and voltage inputs, also achieving an accuracy of 99.542%.

While these approaches showcase notable successes, their effectiveness is contingent upon the availability of infrastructure, monitoring data, and facilities, posing a challenge in less-accessible developing countries like Nigeria [25-28].

Geographic Information Systems (GIS) Mapping Based Solutions

The incorporation of geospatial mapping introduces an additional layer of sophistication and efficiency to outage management. GIS has emerged as an indispensable tool in the realm of power outage management, providing spatial intelligence and analytical capabilities. Notably, the field of GIS remains relatively uncommon, resulting in limited research and implementation for outage management in Nigeria [29].

Studies by [30-31] underscore the potency of GIS in mapping and analysing power systems. The precision afforded by GIS, coupled with mapping and GPS data, facilitates an accurate evaluation of outage locations, including the timing and geographical specifics of occurrences. The integration of GIS with other systems, such as Outage Management Systems (OMS) and Distribution Management Systems (DMS), amplifies the overall efficiency of outage response and resolution.

[32] emphasises the pivotal role of integrating GIS with OMS and DMS for a comprehensive outage analysis. They introduce an algorithm for reporting outages in electricity distribution systems, accentuating GIS's advantages in establishing a geometric network structure for precise analysis.

Furthermore, [33] addresses the implementation of the Integrated Distribution Management System (IDMS) in Active Distribution Networks (ADNs). While emphasising the importance of incorporating advanced technologies like machine learning for fault analysis, optimization, and coordination of distributed energy resources (DERs), the paper underscores the indispensable role of Geographic Information System (GIS) in enabling instantaneous visualisation and analysis of distribution networks. Importantly, the integration of GIS and machine learning, not one at the expense of the other, is deemed crucial for enhancing efficiency, fault detection, and decision-making processes within ADNs.

It's noteworthy that while GIS integration presents substantial benefits, challenges such as the scarcity of comprehensive developers, security concerns, and high-performance hardware requirements are acknowledged.

Gaps in Knowledge

Despite significant progress in outage prediction and management, several knowledge gaps persist in current literature. One prominent gap is the limited integration of GIS technologies into predictive modelling frameworks, particularly within the Nigerian context. While GIS offers unparalleled spatial analysis capabilities, its underutilization hampers comprehensive understanding and visualisation of outage patterns, hindering effective decision-making and response efforts.

Furthermore, existing research often relies on retrospective analyses of outage data, overlooking proactive measures for outage prevention and mitigation. There is a notable absence of predictive modelling frameworks that leverage machine learning algorithms to forecast outage occurrences accurately, reflecting a critical need for innovative approaches to outage management.

Moreover, the scarcity of comprehensive and reliable datasets, especially in regions like Nigeria, poses significant challenges to developing robust predictive models. Historical outage data often contain missing or incomplete information, compromising the accuracy and reliability of predictive algorithms and hindering efforts to develop proactive outage management strategies.

Theoretical Framework

The theoretical framework guiding this project draws upon interdisciplinary principles from electrical engineering, data science, and spatial analysis.

The fundamental principle in electrical engineering that governs how power systems operate is given by Ohm’s law, which is:

$$( ) = \quad ( ) * \quad ( ) \; – (2.1)$$

During power outages or faults in the system, Ohm's Law helps in analysing the behaviour of electrical circuits and networks. By examining voltage, current, and resistance values, engineers identify the location and nature of faults such as overcurrents and short circuits, facilitating timely repairs and restoration of service.

Power system dynamics theory explores the response of electrical networks to external disturbances, including weather-related events.
The mathematical representation of load flow analysis [34] is given by:

$$, \; = \; + \qquad – (2.2)$$

$$, \; = \; + \qquad – (2.3)$$

By calculating real and reactive power losses, specific components or sections of the power grid where power losses are occurring are identified. High losses may indicate inefficient equipment, overloading, or other operational issues that could contribute to power outages. Also, voltage regulation [35] given by:

$$(\%) = \frac{- \quad - \quad -}{\quad} * 100 – (2.4)$$

plays a crucial role in maintaining system stability, especially during transient and dynamic conditions. Fluctuations in load demand, changes in generation output, and other disturbances can impact system voltage and consequently cause outages.
Concepts from machine learning, such as regression analysis, decision trees, and ensemble methods, serve as guiding principles for developing predictive models for outage forecasting.

Regression analysis is a statistical method used to model the relationship between one or more independent variables (features) and a dependent variable (target).
It can be represented thus:

$$= \beta_0 + \beta_1 \, {}_1 + \beta_2 \, {}_2 + ... + \beta \qquad + \varepsilon \; –(2.5)$$

Here, represents the outage occurrence, are the input features e.g weather conditions, system conditions, etc. which determine , and β are the coefficients or weights associated with the input features and ε is the error term.

While traditionally associated with continuous variables, regression analysis can be adapted for binary classification tasks in outage prediction. By modelling outage occurrences as binary outcomes (outage vs. no outage), regression techniques such as logistic regression provide insights into the probability of outage events based on various input features like weather conditions, historical outage data, and system parameters.

Decision trees are a non-parametric supervised learning method used for classification and regression tasks. They partition the feature space based on feature values, aiming to maximise information gain at each split. Decision trees are interpretable and robust against outliers.

In decision trees, the splits are based on conditions such as $< h \quad h$ or $\geq h \quad h$ and provide insights into the nonlinear relationships between input features and outage occurrences. By identifying key decision points, utilities can proactively address risk factors and implement targeted interventions to mitigate outage risks in vulnerable areas.

Ensemble methods combine multiple models to improve predictive performance and generalisation. Random Forests, a popular ensemble method, aggregate predictions from multiple decision trees. For classification, the mode of the predicted classes [36] is computed as

$$\hat{} \; (\hat{}_1, \hat{}_2, \ldots, \hat{}) \quad - (2.6)$$

where $\hat{}$ = final predicted outcome, n = number of trees

In outage prediction, Random Forests enhance predictive accuracy by considering diverse perspectives captured by individual decision trees. By averaging or voting the predictions of constituent models, ensemble methods mitigate overfitting and improve model robustness

By leveraging machine learning algorithms and data-driven approaches, the project can extract valuable insights from diverse datasets and improve the accuracy and reliability of outage predictions.

Incorporating theories from spatial analysis and GIS, including spatial autocorrelation and network analysis, facilitates the visualisation and analysis of outage data in geographical contexts.

Spatial autocorrelation assesses the degree to which outage occurrences exhibit spatial clustering [37]. This is expressed through the Moran's I formula:

$$= \frac{}{} \cdot \frac{\sum\sum (\; - \bar{}\;)(\; - \bar{}\;)}{\sum (\; - \bar{}\;)^2} \quad - (2.7)$$

Where = number of spatial units, represents the spatial weight matrix, and are values of the variables at locations and and $\bar{}$ is the mean of the variable.

It evaluates whether outages tend to cluster in certain geographic areas or if there is randomness in their distribution. Understanding spatial autocorrelation is essential for identifying hotspots or areas with higher outage frequency, which can inform targeted interventions and resource allocation strategies.

Network analysis involves the study of the interconnectedness of outage-affected areas and check how disruptions in one area may impact neighbouring regions. It can be represented by graph theory metrics.

The degree centrality ( ) of a node in a network [38] can be calculated as:

$$= \frac{\qquad}{-1} \quad – (2.8)$$

Integrating spatial analysis techniques with machine learning algorithms enables a comprehensive understanding of outage patterns and spatial relationships. This integration empowers more informed decision-making and resource allocation during outage response efforts.

By synthesising principles from diverse disciplines, the project endeavours to develop a comprehensive theoretical framework and methodology. This approach aims to enhance power outage management, by integrating predictive modelling, machine learning, and spatial analysis techniques. The goal is to mitigate outage impacts and improve system reliability and resilience.

CHAPTER THREE

METHODOLOGY

Research Materials

The materials and tools utilised in this project include various data sources, software applications and programming languages. The primary materials used are:

Data Sources

1. **Ikeja Electric Fault Logs Data**

   This dataset comprises detailed records of power outage incidents within the operational jurisdiction of Ikeja Electric Distribution Company (IKEDC) in Lagos, Nigeria.
   It includes information such as the date of the outage, the branch or area affected, the specific location, feeder type, fault category, and areas impacted.
   The data covers outage incidents from October 2, 2021, to December 2, 2023.

2. **Visual Crossing Weather Data**

   Visual Crossing is a leading provider of weather data and enterprise analysis tools to data scientists, business analysts, professionals, and academics. Visual Crossing Weather is a very easy-to-use and low-cost source for historical and forecast weather data.

   This dataset contains hourly meteorological observations captured across Lagos, Nigeria.
   It includes a wide range of weather parameters such as temperature (minimum, maximum, average), precipitation (probability, type), wind speed, direction, solar radiation, and cloud cover.
   The weather data spans from October 1, 2021, to December 4, 2023, providing detailed insights into weather patterns and trends over time.

3. **The Nigeria Electricity System Operator (SO) Data**

   System Operation (SO) being a semi-autonomous sector under the Transmission Company of Nigeria (TCN) is responsible for operating the transmission system in a safe and reliable manner. This data encompasses daily energy generation metrics and system operational parameters pertinent to Nigeria's electricity grid. This dataset includes information such as the date of observation, peak generation values, daily energy generation figures, lowest energy generation levels, energy dispatched daily at 6:00

am, highest system frequency recorded, lowest system frequency recorded, highest voltage recorded, and lowest voltage recorded.

NESO data is crucial for understanding the underlying factors contributing to power outages and grid disturbances. By analysing this data, insights can be gained into the relationship between energy generation patterns, system operation, voltage levels, frequency stability and the occurrence of electricity outages. Additionally, this information is valuable for developing predictive models and implementing measures to enhance the reliability and efficiency of Nigeria's electricity supply network.

4. **OpenStreetMap (OSM) Data**

OpenStreetMap (OSM) is a collaborative project that provides freely accessible geospatial data. The dataset includes vector features such as roads, buildings, points of interest, administrative boundaries, and other geographic elements.
It offers comprehensive spatial coverage of Nigeria, enabling detailed mapping and analysis for various applications, including geospatial visualisation and spatial analysis.

Structure, Granularity, Scope, Temporality and Faithfulness (SGSTF) of the Datasets

1. **Ikeja Electric Fault Logs Data**

The Ikeja Electric fault log data is structured in a tabular format, providing detailed information about individual fault incidents, including the date, location, fault description, fault type, and affected areas. This organisation allows for easy querying and analysis, making it applicable for establishing fault patterns and trends. With granularity at the level of individual fault incidents, the dataset enables analysis of outage frequency and distribution across different branches and locations within the Ikeja Electric distribution network.

This level of detail ensures its applicability for identifying recurring patterns and hotspots of outage activity. Covering power outage incidents within the operational jurisdiction of Ikeja Electric Distribution Company (IKEDC) in Lagos, Nigeria, the dataset offers insights into common causes of outages, such as equipment failures and technical faults. This broad scope makes it applicable for understanding outage dynamics in various operational contexts. The dataset includes historical records of power outage incidents from October 2, 2021, to December 2, 2023, ensuring its relevance for analysing outage trends over time. Systematic collection by Ikeja Electric Distribution Company ensures the accuracy and completeness of the data, enhancing its applicability for analysis and modelling across different distribution companies and regions.

2. **Visual Crossing Weather Data**

The Visual Crossing weather data is organised in a tabular format where rows represent daily intervals and columns capture various weather attributes. Each row includes details on temperature (minimum, maximum, average), precipitation (probability, type), wind speed, cloud cover, solar radiation, and other meteorological variables. This structured format allows for clear interpretation and analysis of weather patterns and trends over time.

The dataset provides daily weather observations, offering detailed insights into temperature fluctuations, precipitation levels, wind patterns, and other atmospheric conditions. It covers meteorological data recorded across Lagos, Nigeria, from October 1, 2021, to December 4, 2023. The high-temporal resolution of the dataset enables detailed temporal analysis, facilitating the identification of seasonal patterns, weather trends, and anomalies that may impact power system reliability and outage occurrences. The data's accuracy and reliability are ensured by Visual Crossing, a reputable provider, which conducts quality assurance and validation checks to maintain the dataset's integrity for analysis and modelling.

3. **NESO (Nigeria Electricity System Operator) Data**

The NESO data is structured in a tabular format where rows represent individual observations or daily snapshots of the electricity system, with columns capturing metrics such as date, peak generation, daily energy generation, lowest energy generation, energy dispatched at 6:00 am, system frequency, and voltage levels. Each row provides a chronological sequence of observations, enabling easy querying and analysis of different parameters related to energy generation and system operation. The data granularity is at the daily level, offering a broad overview of daily trends and patterns but lacking the detail for real-time or intra-day analysis.

It covers various aspects of electricity generation and system operation in Nigeria, including energy generation capacity, operational performance, and grid stability. This scope allows for monitoring and analysis of key parameters influencing the reliability, efficiency, and resilience of the electricity supply system. The dataset is temporally aligned, with each observation time-stamped by date, providing a historical perspective that supports trend analysis, forecasting, and identification of patterns or anomalies affecting grid reliability and stability. The data’s reliability is ensured as it is sourced from the

System Operation (SO) sector under TCN, which is responsible for managing Nigeria's transmission network and is known for its credibility and expertise in operating the transmission infrastructure.

4. **OpenStreetMap (OSM) Data**

The OpenStreetMap (OSM) data is structured as a collection of nodes, ways, and relations, representing geographic features such as roads, buildings, landmarks, and natural elements like rivers and forests. Nodes are individual points defined by latitude and longitude coordinates, ways are ordered lists of nodes representing linear features, and relations group nodes, ways, and other relations to represent complex features.

The data offers fine granularity, providing detailed information about geographic features at various scales, allowing for precise mapping and analysis. OSM's global scope includes diverse geographic features like roads, buildings, land use, and transportation networks, covering urban, rural, and remote areas. The dataset is dynamic and continuously updated, reflecting real-world changes and contributions from a global user community. Frequent updates, along with temporal information indicating the date of modifications, ensure the data's relevance. The fidelity and accuracy of OSM data are maintained through community-based validation and peer review processes, which help ensure data quality for GIS applications and mapping purposes.

Software and Programming Languages

Python is a high-level, interpreted programming language celebrated for its simplicity, readability, and versatility, making it a preferred choice for both beginners and experienced developers. Its extensive ecosystem, encompassing a wide range of libraries and frameworks, has made Python indispensable in fields like data analysis, machine learning, and geospatial visualisations. In this project, Python was selected as the primary programming language due to its flexibility and powerful capabilities, serving as the foundation for data preprocessing, machine learning model development, and geospatial mapping.

To achieve the project's objectives, several key Python libraries were used. Pandas and NumPy were employed for data manipulation and numerical operations, enabling efficient handling of large datasets and complex calculations. Matplotlib and Seaborn were essential for data visualisation, allowing the creation of detailed charts and graphs that helped uncover underlying patterns in the data. Scikit-learn played a pivotal role in the development and evaluation of machine learning models, offering comprehensive tools for tasks such as classification, regression, and clustering. GeoPandas extended Pandas' functionality to handle geospatial data, facilitating seamless integration of geographic information into the project's analytical workflows. Additionally, FastAPI, a modern and high-performance Python web framework, was used to develop the web-based application designed for monitoring electricity outages. This framework supported the creation of a fast, scalable backend, crucial for real-time data processing and user interaction. The user interface, developed using HTML, CSS, and JavaScript, ensured the platform was both visually appealing and easy to navigate, enhancing the overall user experience.

In terms of development environments, Jupyter Notebook and Visual Studio Code were the primary Integrated Development Environments (IDEs) utilised throughout the project. Jupyter Notebook provided an interactive computing environment that allowed for the creation and sharing of documents containing live code, visualisations, and narrative text. This interactive nature was particularly advantageous, enabling a streamlined process for documenting code, annotating findings, and visualising results in a single, cohesive interface. It was ideal for iterative data analysis, model development, and thorough documentation of research methodologies.

Visual Studio Code, a lightweight, cross-platform source code editor by Microsoft, complemented Jupyter Notebook by serving as the primary environment for writing more complex code, especially for the web application. With built-in support for multiple languages and frameworks like Python and JavaScript. Visual Studio Code provided features such as syntax highlighting, code completion, and integrated debugging tools. These made it well-suited for both front-end and back-end development tasks, offering a customizable and efficient workspace for managing the project's various components. Its versatility ensured a smooth and productive development process, accommodating the project's diverse needs from Python development to web design.

Hardware Used

The MacBook Pro 2020 laptop was the central computing platform for this project's extensive data analysis, model development, web development, and geospatial visualisation tasks. Its robust hardware capabilities ensured smooth performance and efficient processing of the demanding computational tasks required for the project. Equipped with 500GB of storage space, the laptop provided ample memory for storing datasets, models, and all other project-related files, ensuring that storage limitations never hindered progress.

At the heart of the MacBook Pro is the Apple M1 chip, a cutting-edge processor that integrates an 8-core CPU, featuring 4 performance cores and 4 efficiency cores. This architecture allowed the laptop to handle complex computations and multitasking with ease, making it ideal for the project's varied demands, from running machine learning algorithms to rendering geospatial maps. The 8-core GPU further enhanced the laptop's capability, enabling high-performance graphics processing essential for the detailed visualisations and interactive elements of the web application. Together, these hardware specifications made the MacBook Pro 2020 a powerful and reliable tool, perfectly suited to the technical requirements of the project.

Methods Adopted

The methods for this study involved key steps aimed at developing an outage prediction model for electricity outages and geospatial mapping as illustrated in the flowchart below:

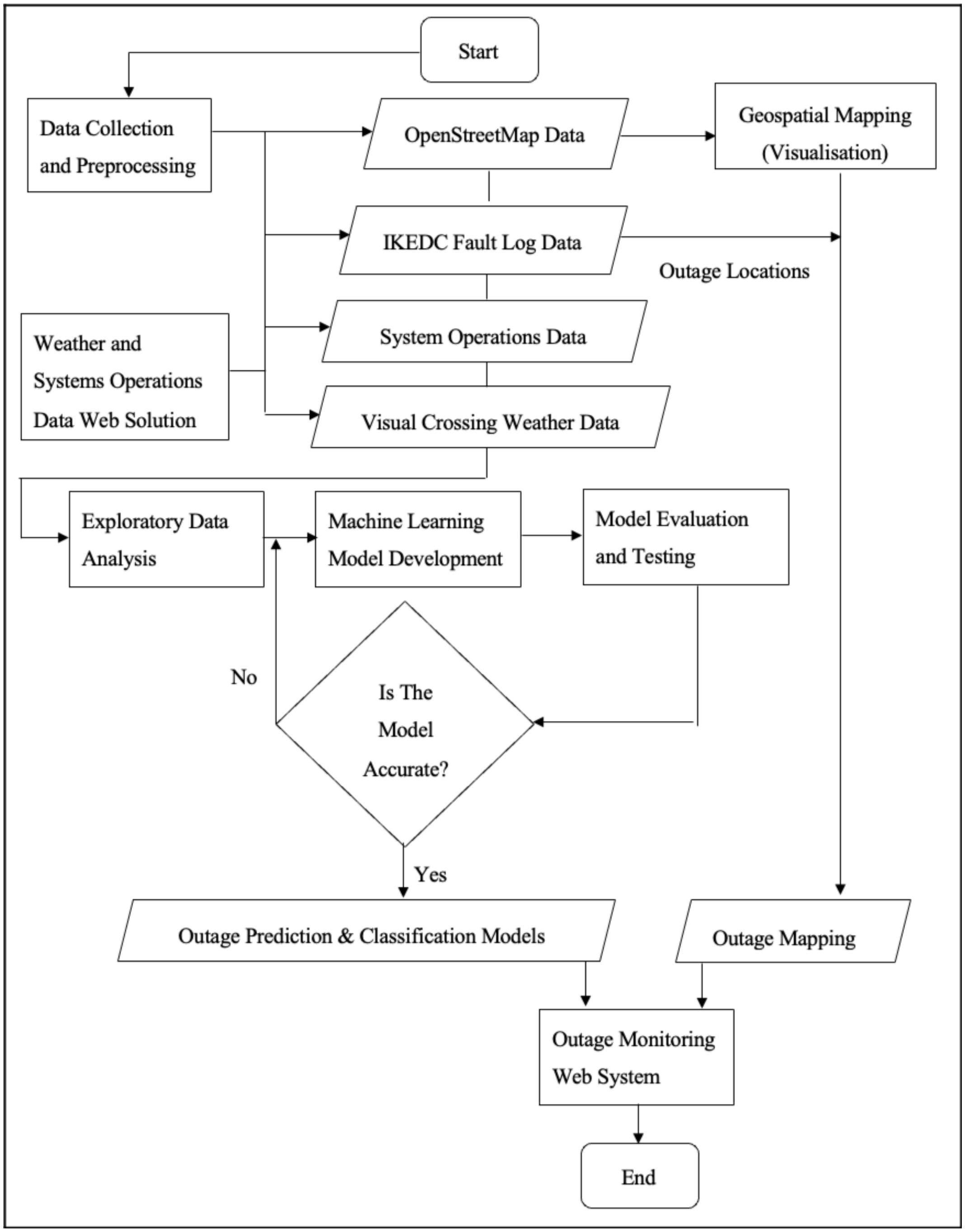


Figure 3.1: Flowchart Outlining the Methods Adopted in the Software Development Life Cycle (SDLC) of the Project.

Data Collection and Preprocessing

1. **IKEDC Fault Logs Data**

   The Ikeja Electric's fault log data was manually inputted, initially comprising details like the fault date, branch, location, fault description, and affected areas. However, due to the often unclear fault descriptions, a separate column for fault type was established to enhance clarity regarding the fault nature.

   Additionally, efforts were made to classify faults based on feeder types (11KV, 33KV, 132KV), though this categorisation was not always clear from the fault descriptions directly. Categorical variables such as fault type were appropriately encoded to ensure compatibility with machine learning models. This approach aimed to tackle the challenge of manually interpreting fault types from ambiguous descriptions using machine learning techniques, which were deployed as part of an outage classification web system.

2. **Visual Crossing Weather Data**

   The Visual Crossing weather data was integrated with the fault logs data based on the corresponding dates and IKEDC branch. Dates were used to filter the temporal extent, matching the fault logs' rows with the weather data rows, while the branch helped determine the location for obtaining the necessary weather parameters for that particular date. Data alignment and standardisation were performed to ensure consistency and compatibility between the datasets. This data is utilised within the outage prediction web system, allowing users to specify dates and access detailed weather information to enhance predictive models, thereby improving infrastructure resilience.

3. **The Nigeria Electricity System Operator (SO) Data**

   The SO data was processed to extract relevant features such as peak generation, daily energy generation, and system frequency. Missing information in this data was handled using a developed random forest model for data imputation. This data is also employed in the web system to predict outages, enhancing the accuracy of the models by integrating critical operational metrics.

Exploratory Data Analysis (EDA)

1. **Statistical Analysis**

   Descriptive statistics were computed to summarise key variables such as mean, standard deviation, and range, providing insights into the central tendency and dispersion of the data. The distribution of outage frequencies across different categories, including fault types and locations, was analysed to identify prevalent outage patterns. This helped in understanding which types of faults and locations were most prone to outages.

Correlation analysis was then performed to explore relationships between various variables. For example, the correlation between weather conditions, such as temperature, wind speed, and sea level pressure, and outage occurrences was examined. Additionally, the correlation between the transmission system's operational data, like energy generated and energy sent out, was analysed to confirm that the majority of outages occurred on the distribution side. Understanding these correlations is crucial for identifying potential drivers of power outages.

To further analyse the data, distribution plots were created, including time-series plots that helped in identifying temporal patterns and distributions of outages, energy generation, and weather parameters. These plots were instrumental in revealing seasonal trends and periodic fluctuations. The frequency distribution of categorical variables, such as fault types and locations, was also examined using bar charts, allowing for the identification of dominant categories, rare events, and specific locations that were particularly susceptible to outages.

2. **Spatial Analysis**

Spatial analysis in this project began with examining spatial autocorrelation, which refers to the tendency of similar events to be geographically close to each other. Outage maps were used to assess the degree to which similar outage events occurred in proximity within the distribution network. By analysing spatial data that included timestamps, locations, and descriptions of outages, geospatial techniques helped determine whether areas experiencing outages were clustered together or randomly spread out. This analysis was crucial for identifying specific regions that consistently faced more outages, providing insights into potential underlying causes.

Clustering analysis further explored the geographic distribution of outage clusters, revealing common causes and the spatial extent of these events. Outage clusters were visualised on maps using colour-coded markers that indicated the density and severity of outages, making it easier to identify frequent and severe outage hotspots. This visualisation was invaluable for electrical engineers and utility managers, allowing them to prioritise areas requiring maintenance and upgrades. Additionally, the map provided a clear understanding of the spatial distribution of different types of faults, such as transformer and feeder faults, and their geographic occurrence. This information was essential for strategic planning and resource allocation, ultimately aiming to enhance the overall reliability and resilience of the power distribution network.

Machine Learning Model Development

1. **Outage Prediction Model (OPM)**

The Outage Prediction Model (OPM) focused on predicting power outages a day (24 hours) in advance using historical data. For this, three machine learning algorithms were evaluated: Logistic Regression, Random Forest, and Neural Networks. Logistic regression's architecture revolves around its ability to predict binary outcomes, such as whether an outage will occur based on various input features. The model forms a linear combination of these features, weighted by coefficients that are learned during training. This combination, along with a bias term, forms the basis of a logistic function (sigmoid), which transforms the linear output into a probability score between 0 and 1. This score represents the likelihood of an outage. Once trained, the model can predict the probability of outages for new instances by applying the learned coefficients to their respective features. This prediction capability is crucial for understanding and mitigating outage risks in various operational contexts.

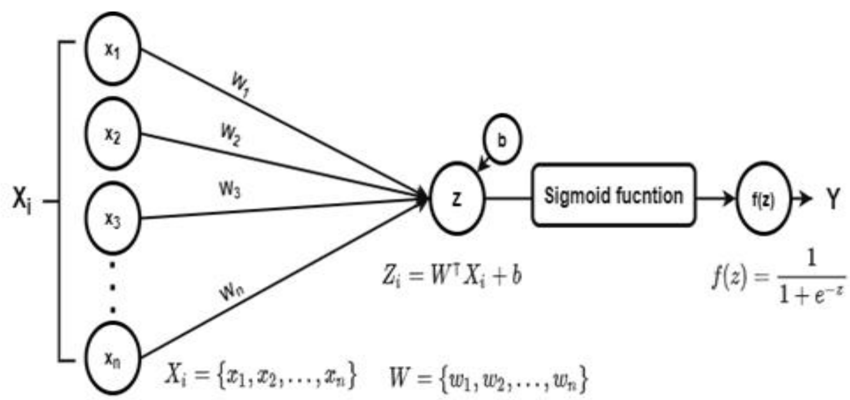


Figure 3.2: Logistic Regression Model Architecture [39]

Random forests are an ensemble learning technique that constructs multiple decision trees and combines their predictions to improve accuracy and reduce overfitting. Each tree in the forest is trained on a bootstrap sample of the dataset, and feature randomness is introduced during tree construction to promote diversity among the trees.

A random forest consists of a collection of decision trees, where each tree is trained independently on a random subset of the training data and a random subset of features. During prediction, the output of each tree is aggregated (e.g., through averaging for regression or voting for classification) to produce the final prediction.

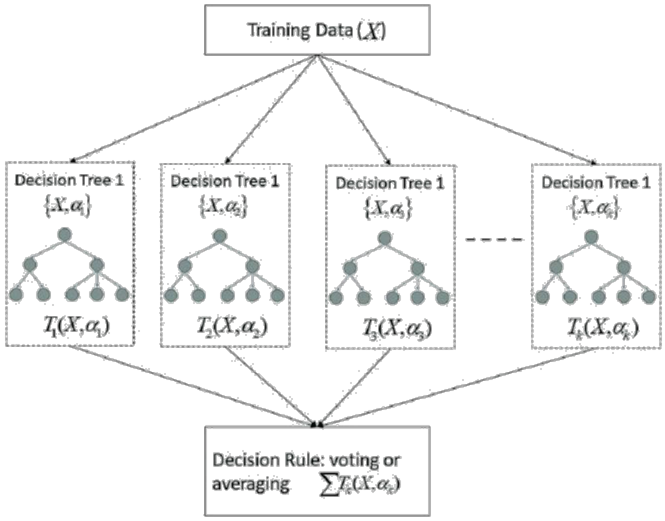


Figure 3.3: Random Forest Model Architecture [40]

Neural networks are a class of machine learning models inspired by the structure and function of biological neural networks. They consist of interconnected nodes organised into layers, including an input layer, one or more hidden layers, and an output layer. Each node performs a weighted sum of its inputs, followed by a non-linear activation function.

The architecture of a neural network can vary widely depending on factors such as the number of layers, the number of nodes per layer, and the choice of activation functions. Common architectures include feedforward neural networks, convolutional neural networks (CNNs) for spatial data, and recurrent neural networks (RNNs) for sequential data. Training involves forward propagation of input data through the network, followed by backpropagation of errors to adjust the network's parameters (weights and biases) using optimization techniques like gradient descent.

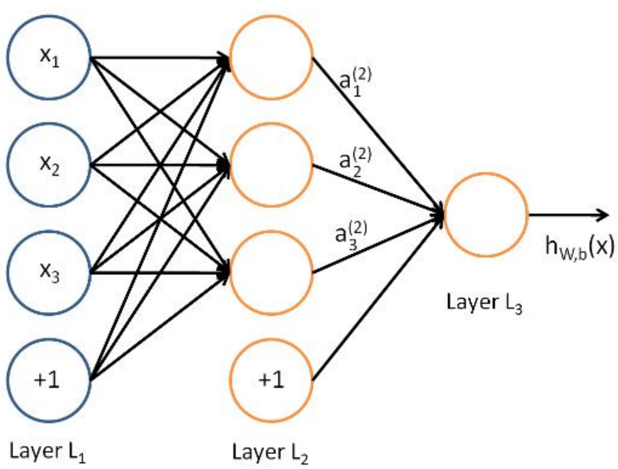


Figure 3.4: The Three-Layer Neural Network [41]

Logistic Regression was initially selected for its simplicity and interpretability, but it showed lower performance compared to the other models. Random Forest demonstrated superior predictive capabilities, particularly in distinguishing between outage and non-outage instances due to its robust nature. Although Neural Networks have the potential to capture more intricate data patterns, they were not pursued further due to the extensive fine-tuning and limited time constraint within the project deadlines.

2. **Outage Classification Model**

To address the issue of ambiguous fault descriptions in the Ikeja Electric Fault Logs, an outage classification model was to be developed. This model aimed to categorise faults into distinct types based on textual descriptions using machine learning techniques, particularly Natural Language Processing (NLP) and classification algorithms.

NLP involves the interaction between computers and humans using natural language. It enables computers to understand, interpret, and generate human language. Classification algorithms, on the other hand, are machine learning techniques used to categorise data into predefined classes or categories based on input features. The fault classification model employs a pipeline consisting of two main components: a TF-IDF vectorizer and a classification model of choice, which could be either a decision tree or random forest model.

TF-IDF Vectorizer: This component transforms text data into numerical vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) method. TF-IDF measures the importance of a word in a document relative to a collection of documents, helping to identify key terms and their significance in fault descriptions.

Classification Model: The classification model serves as the main classifier responsible for categorising fault descriptions into predefined fault classes. The fault classes are: Breaker, Feeder, High-Voltage (Distribution) System, Human, Customer Notification, Planned, Substation, Transmission, and Transformer faults. After rigorous iterations and comparative analyses, the random forest emerged as the model of choice due to its superior performance over alternatives like logistic regression and Support Vector Classifier (SVC).

The model underwent testing and validation to ensure accurate fault classification, with appropriate metrics used to evaluate its effectiveness.

Model Evaluation and Testing

Performance metrics were employed to rigorously evaluate the effectiveness of the developed models in predicting outages and handling missing data within the system operations. These metrics provide insights into the models' accuracy, robustness, and reliability.

For the outage prediction model, several key metrics were employed: Precision, which indicates the proportion of correctly predicted positive instances (outages) among all instances predicted as positive. Recall, which measures the proportion of correctly predicted positive instances (outages) among all actual positive instances. F1-score, a harmonic mean of precision and recall that provides a balanced measure between the two metrics.

Precision: Measures the proportion of correctly predicted instances (true positives) of a specific class (outage or non-outage) out of all instances predicted as that class (true positives or false positives).

= + - (3.1)

Recall: Measures the proportion of correctly predicted instances (true positives) of a specific class (outage or non-outage) out of all actual instances of that class (true positives or false negatives).

= + - (3.2)

Accuracy or F1 Score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance for each class.

1= 2* + - (3.3)

Area Under ROC Curve (AUC), which evaluates the model's ability to distinguish between positive (outage) and negative (non-outage) instances, with higher values indicating better performance.

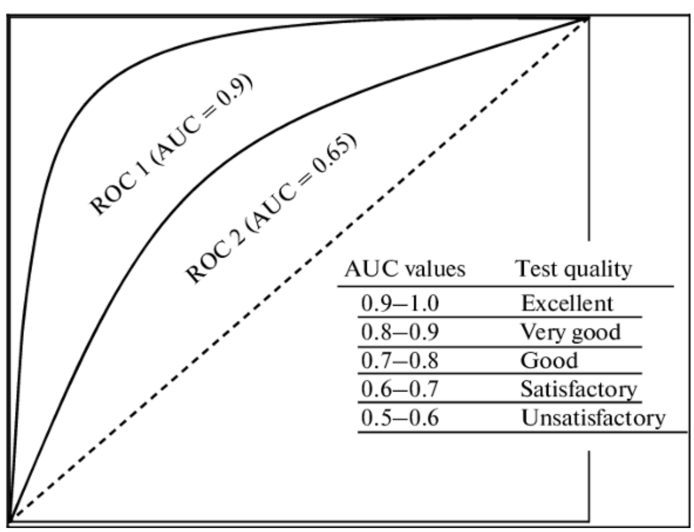


Figure 3.5: Area Under the ROC Curve (AUC) Interpretation of Results

In the system operations regression model, the Mean Absolute Deviation (MAD) served as a metric to assess the accuracy of predicting missing data. MAD measures the average absolute difference between actual and predicted values, offering insights into the model's precision in handling data discrepancies.

Furthermore, for the outage classification model, precision and recall were again utilised to gauge the model's accuracy in identifying different outage types. Additionally, Cohen's Kappa Score and Matthews Correlation Coefficient (MCC) were employed. The earlier indicates good agreement, suggesting that the model's predictions are very reliable while the latter indicates a strong correlation between the predicted and actual classes, taking into account the size of the classes and their balance. The confusion matrix was also utilised to provide a detailed breakdown of the model's performance in classifying fault types, illustrating the types of errors made by the model in its predictions.

Cross-validation techniques such as K-fold cross-validation was applied to ensure robustness and generalisability of the outage related models. Python's GridSearchCV is a popular utility that was used to accomplish this task.

The developed models were tested against data unseen from the model development and validation stage, known as the test set. This was a good measure of estimating the performance of the model on fresh data.

Geospatial Mapping (Visualisation)

Geospatial Mapping played a pivotal role in this project, focusing on transforming raw outage data into insightful visualisations using advanced geographic information system (GIS) techniques. The process began with meticulous data preparation, integrating outage logs, timestamps, locations, and descriptive details for comprehensive analysis.The visualisation techniques employed were tailored to provide clear insights into outage patterns and their spatial distribution.

Outage Heat Map vis. was utilised to depict outage hotspots across different regions of the service area. These colour-coded maps effectively highlighted areas with frequent and severe outages, facilitating targeted interventions and infrastructure upgrades to improve resilience.

Cluster Mapping further dissected outage clusters, revealing their geographic concentration and common causes. By identifying hotspots and comparatively stable areas, this approach enables utility managers and engineers to prioritise maintenance efforts and allocate resources strategically and helps in distinguishing isolated incidents from systemic problems. This tool supports strategic planning for resilient infrastructure development.

Temporal Analysis through Time-Series Snapshot Mapping provided a dynamic view of outage occurrences over time, seeing how they evolve. This chronological representation helped in identifying recurring patterns, assessing the impact of weather events, and refining response strategies to enhance grid resilience. Interactive web-based tools were developed to enable geolocation filtering, allowing stakeholders to zoom into specific branches or localities. This granular level of detail facilitated a deeper understanding of outage dynamics, aiding in rapid response and targeted maintenance.

The outcomes of these geospatial mapping techniques extended beyond visualisation; they supported data-driven decision-making by evaluating power distribution equity and identifying areas requiring immediate attention for infrastructure improvements. This comprehensive approach not only evaluates operational efficiency but also assists in decision making to strengthen the overall resilience of the distribution network.

Outage Monitoring Web-Based System

After successfully developing, evaluating, and testing the outage prediction model and creating an accurate, interactive geospatial visualisation of outages, the next phase involved integrating these components into a comprehensive web application. Additionally, a dedicated webpage was developed to fetch Nigeria-specific System Operations and Weather Data. This data extraction component enables precise date specification and access to detailed weather information, crucial for enhancing the outage prediction model and overall infrastructure resilience. By integrating Nigeria's system operations and weather data, the system enhances the accuracy of outage prediction models. Users can specify dates to access detailed weather information, essential for understanding environmental factors contributing to outages. The integration of real-time and historical weather data significantly improves predictive model performance by incorporating variables such as temperature, humidity, and sea-level pressure. This data integration is instrumental in identifying patterns and trends related to weather-induced outages.

The frontend interface was developed using HTML, CSS, and JavaScript to ensure a user-friendly experience. This interface includes features like date and location pickers, outage data display, predictions data download, and visualisation or mapping tools, facilitating seamless interaction.

For efficient backend functionality, the FastAPI framework was chosen due to its capability in handling asynchronous requests and data processing. It serves as the backbone for integrating predictive models, managing data retrieval processes, and implementing user authentication features.

Deployment on Heroku ensures scalability and accessibility, catering to stakeholders involved in outage management and infrastructure planning.

The integrated system underwent rigorous evaluation and testing, including simulated scenarios, real-time data validation, and user acceptance testing. These processes were crucial in validating the system's performance, accuracy, and reliability, ensuring it meets operational requirements and provides actionable insights for outage management and infrastructure resilience.

CHAPTER FOUR

RESULTS AND DISCUSSION

This section presents and analyses the results obtained from the implementation of the project, which focuses on accurate modelling with GIS techniques to identify unplanned High Impact Low Probability (HILP) outages on the IKEDC Distribution Network. This section validates the effectiveness of the various systems and models developed, including the outage prediction system, outage classification system, integration of system operations and weather data, and the outage mapping system. The discussion interprets these results in the context of the project's objectives and the broader implications for power distribution resilience and outage mitigation. The results are systematically organised to cover the following key areas:

Outage Prediction System: Evaluation of the predictive performance of the Random Forest model used to forecast both planned and unplanned outages.

Outage Classification System: Assessment of the accuracy and effectiveness of the system in categorising fault descriptions from fault logs into predefined fault types.

System Operations and Weather Data Analysis: Analysis of how incorporating system operations and weather data improves outage prediction models and overall network resilience.

Outage Mapping System: Visualisation and analysis of outage data over time, identification of geographic clusters, and implications for targeted interventions and power reliability improvement.

Findings from these areas are discussed in detail, highlighting the significance, implications and any unexpected results. This section also provides practical recommendations for enhancing the resilience of the IKEDC Distribution Network and suggests avenues for future research and development based on the observed results.

Exploratory Data Analysis (EDA) and Findings

Initial study revealed different types of outages occurred in the distribution fault logs, which could be categorised into: Breaker, Feeder, High-Voltage (Distribution) System, Human, Customer Notification, Planned, Substation, Transmission, and Transformer faults, which are described thus:

- **Breaker:** Faults related to circuit breakers, which are devices designed to protect electrical circuits by interrupting current flow in the event of an overload or short circuit.
- **Feeder:** Issues occurring on feeders, which are distribution lines that carry electricity from substations to local distribution points.

- **High-Voltage (Distribution) System:** Faults within the high-voltage distribution network, affecting the transmission of electricity over long distances from power plants to substations.
- **Human:** Outages caused by human error or interference, such as operational mistakes, accidental damage during maintenance activities, vandalism.
- **Customer Notification:** Planned or unplanned outages specifically communicated to customers, typically for maintenance or emergency repairs.
- **Planned:** Scheduled outages for routine maintenance, upgrades, or infrastructure improvements to ensure the reliability and safety of the power network.
- **Substation:** Issues occurring at substations, where voltage levels are adjusted (stepped up or down) for efficient power transmission and distribution.
- **Transmission:** Faults within the transmission system, involving high-voltage lines that carry electricity from power plants to substations across long distances.
- **Transformer:** Issues related to transformers, which are devices that change the voltage level of electricity to make it suitable for transmission and distribution.

The bar chart below summarises the fault occurrences by category:

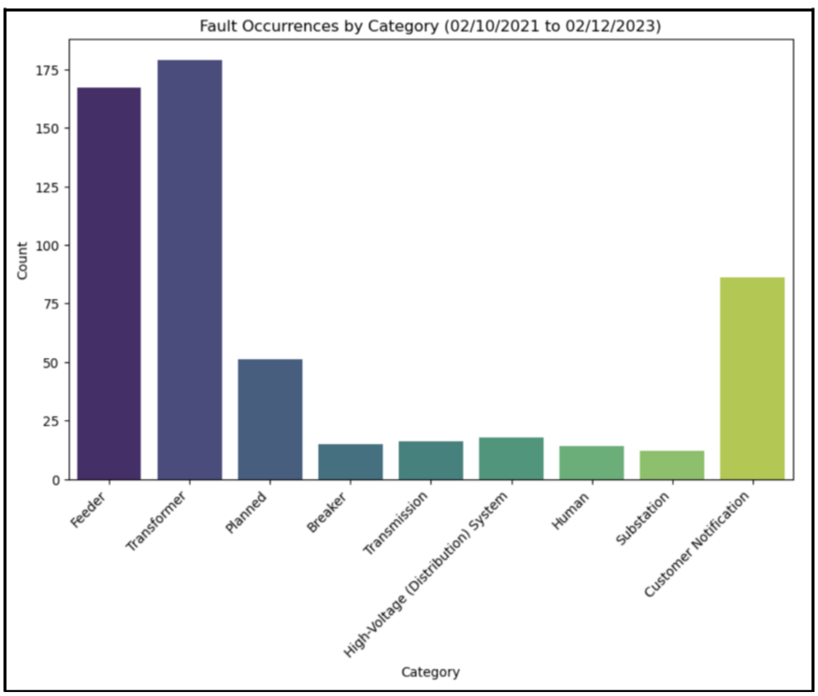


Figure 4.1: Bar Chart Illustrating Outage Distribution by Category

Analysis for the data gathered during the period of 2nd October, 2021 to 2nd December, 2023 revealed that the majority of outages were unplanned, accounting for over 75% of the total. A smaller proportion were planned outages. The exact percentage could not be determined because some fault descriptions, especially those related to customer notification outages (which may or may not be planned), could not be fully interpreted without internal consultation with the distribution company.

Fault logs showed a high frequency of 'Transformer' and 'Feeder' faults (62% of total outages), indicating common areas of failure in the network. 'Human' and 'Substation' faults contributed least to the outage amount. The exact fault type distribution is shown in the table below:

Table 4.1: Distribution of Outage Frequency By Outage Type

Outage Type	**Outage Frequency**
Breaker	15
Customer Notification	86
Feeder	167
High Voltage (Distribution) System	18
Human	14
Planned	51
Substation	12
Transformer	179
Transmission	16

Furthermore, the branches from IKEDC were: Abule Egba, Akowonjo, Ikeja, Shomolu, Oshodi and Ikorodu, with the most faults was Abule Egba, responsible for 28% of the total outages with Iju being the most prevalent location (8% of all). The distribution can be found in the bar chart below:

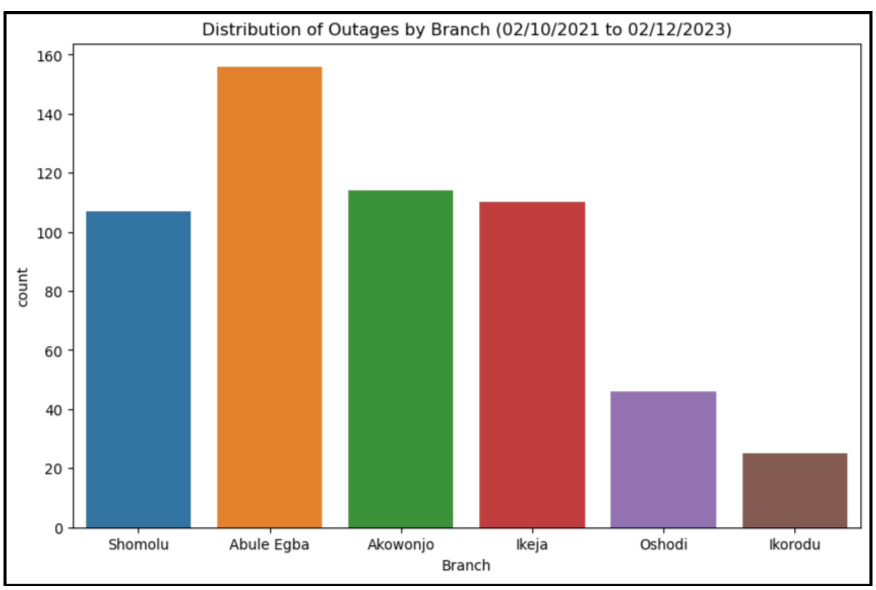


Figure 4.2: Bar Chart Showing the Distribution of Outages by Branch

The Systems Operations (SO) data was analysed to gauge its impact on outages by comparing transmitted energy with generated energy, thereby assessing transmission efficiency. High transmission efficiency indicates that most generated energy reaches its destination with minimal losses, suggesting that distribution side issues, rather than transmission inefficiencies, are the primary cause of outages. Therefore, prioritising the resolution of distribution side faults targets the main source of outages strategically, ensuring interventions are directed where they can most effectively enhance power reliability.

The time-series chart below shows the pattern of energy sent against the energy generated to evaluate efficiency.

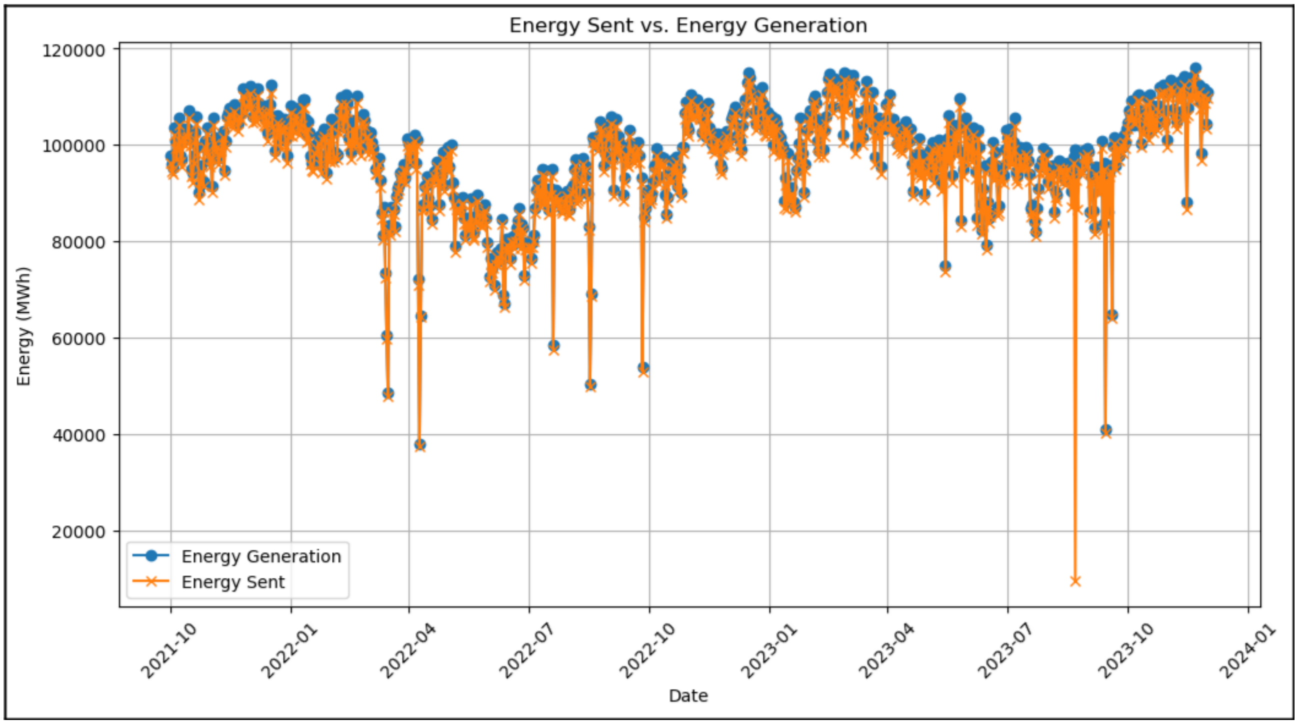


Figure 4.3: Energy Generated vs Energy Sent Illustrating High Transmission Efficiency.

The average efficiency of energy transmission, calculated at 98.52%, indicates that nearly all generated energy successfully reaches its intended destinations with minimal losses.

Low variation in voltage and frequency on the transmission side means that the power supply is stable and consistent, reducing the likelihood of power quality issues and ensuring reliable operation of electrical equipment. This reinforces the focus on addressing distribution side faults as the main source of outages, ensuring that efforts to improve power reliability are directed appropriately.

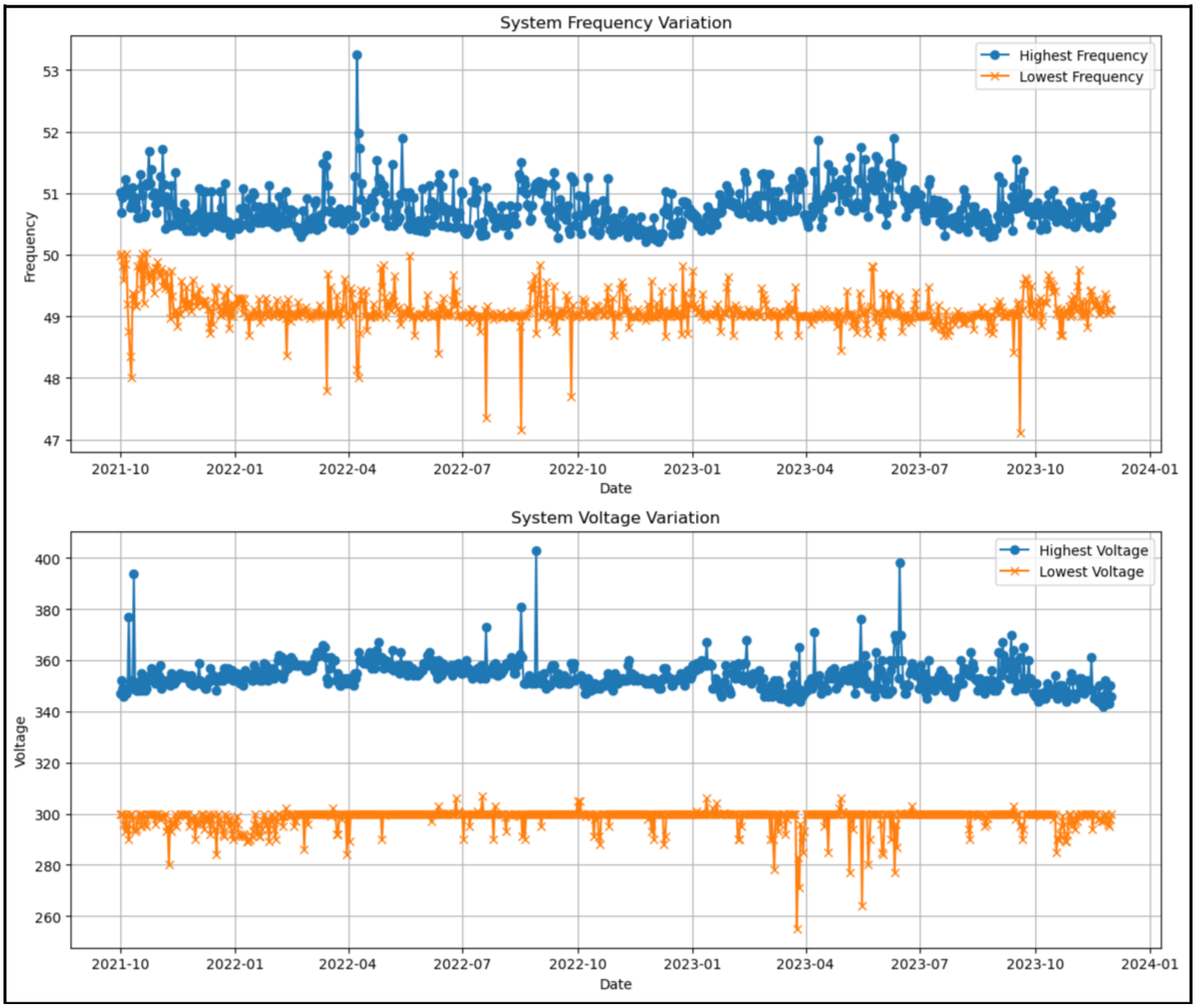


Figure 4.4: Highest and Lowest System Voltages and Frequencies Showing Low Variation.

Meteorological conditions were also evaluated using weather data from Visual Crossing to analyse how outages evolved over time in relation to these conditions. Parameters assessed included temperature and sea level pressure.

Analysis of the data revealed a significant correlation between temperature extremes and the frequency of outages. Specifically, outages tended to increase when temperatures reached extreme highs or lows. This observation aligns with the understanding that extreme temperatures can place additional stress on electrical infrastructure, leading to a higher likelihood of faults and outages.

A noteworthy trend was observed with sea level pressure. The data indicated that as sea level pressure dropped, the number of outages increased. Lower sea level pressure is often associated with stormy weather conditions, which can contribute to electrical faults due to factors such as strong winds, lightning, and heavy rainfall.

Both trends can be observed in the time-series figure plot below:

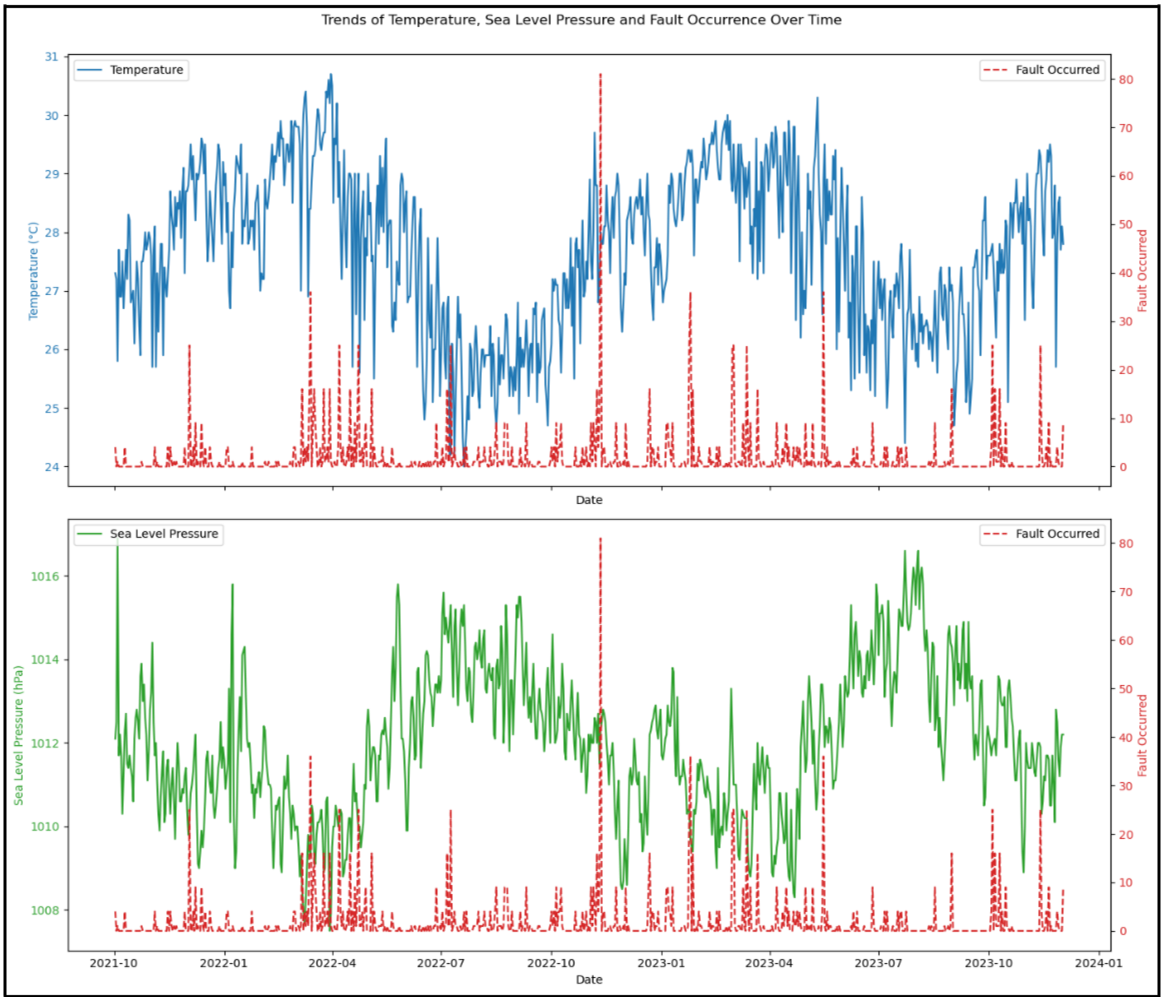


Figure 4.5: Plot of Temperature, Sea Level Pressure and Fault Occurrence Over Time

Systems Operations and Weather Data Solution

1. Model Performance and Evaluation

For predicting the unavailable system operations data from the Nigeria Electricity System Operator (NESO), three models were evaluated: Logistic Regression, Random Forest Regression, Support Vector Regression (SVR) and XGBoost.

The model's performance was evaluated using the Mean Absolute Error (MAE) metric, which measures the average magnitude of errors in the predictions. The following table summarises the MAE results for various target variables:

Table 4.2: Mean Absolute Error (MAE) for the System Operations Target Variables

Target Variable	**Mean Absolute Error (MAE)**
Peak Generation	92.57
Daily Energy Generation	2355.36
Lowest Energy Generation	188.27
Daily Energy Sent	2702.80
6:00 Generation	152.97
Highest System Frequency	0.13
Lowest System Frequency	0.12
Highest Voltage Recorded	2.35
Lowest Voltage Recorded	1.42

The low MAE values relative to the range of values for each target variable indicate that the model achieved high accuracy in predicting essential system operations data. This accuracy is crucial for maintaining reliable power system operations and enabling effective outage management.

2. Integration into the Web Application

The developed model was seamlessly integrated into the web application, providing users with an intuitive platform for extracting and analysing Nigeria's system operations and weather data as seen below:

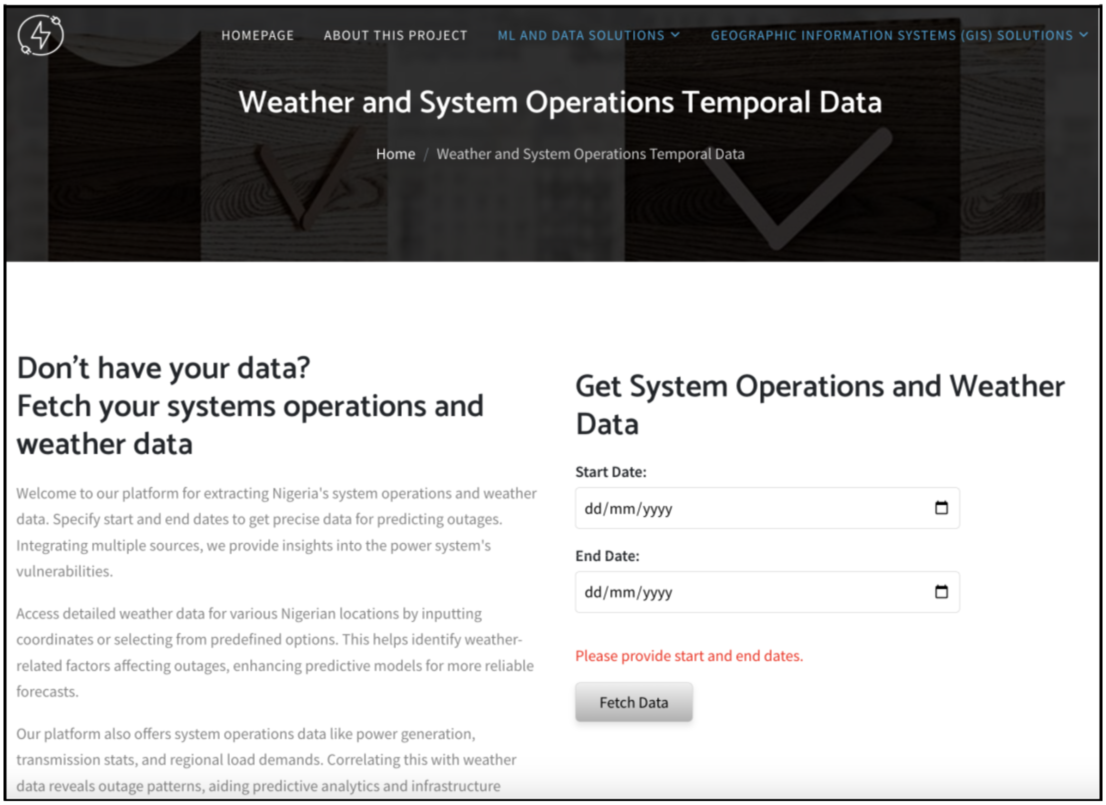


Figure 4.6: First Look at the System Operations and Weather Data Web Solution

When selecting start and end dates, the webpage automatically prompts for locations corresponding to each date, such as location 1 for the first date and subsequent selections for subsequent dates. The selected locations are used to fetch weather information corresponding to each date.

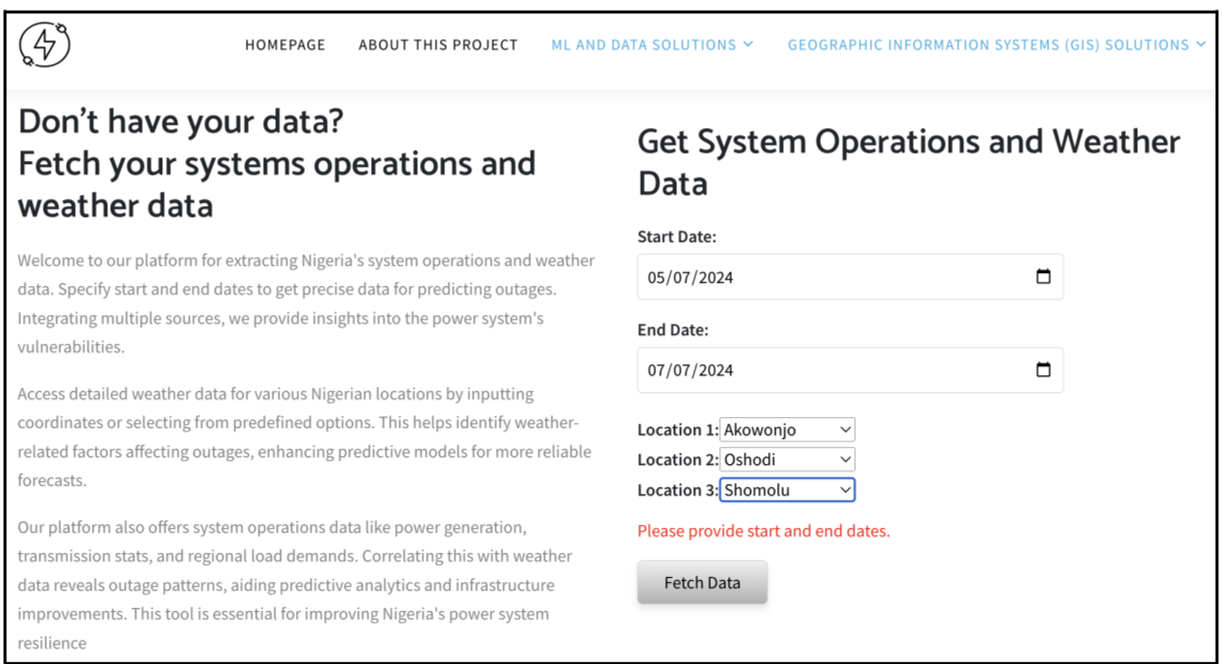
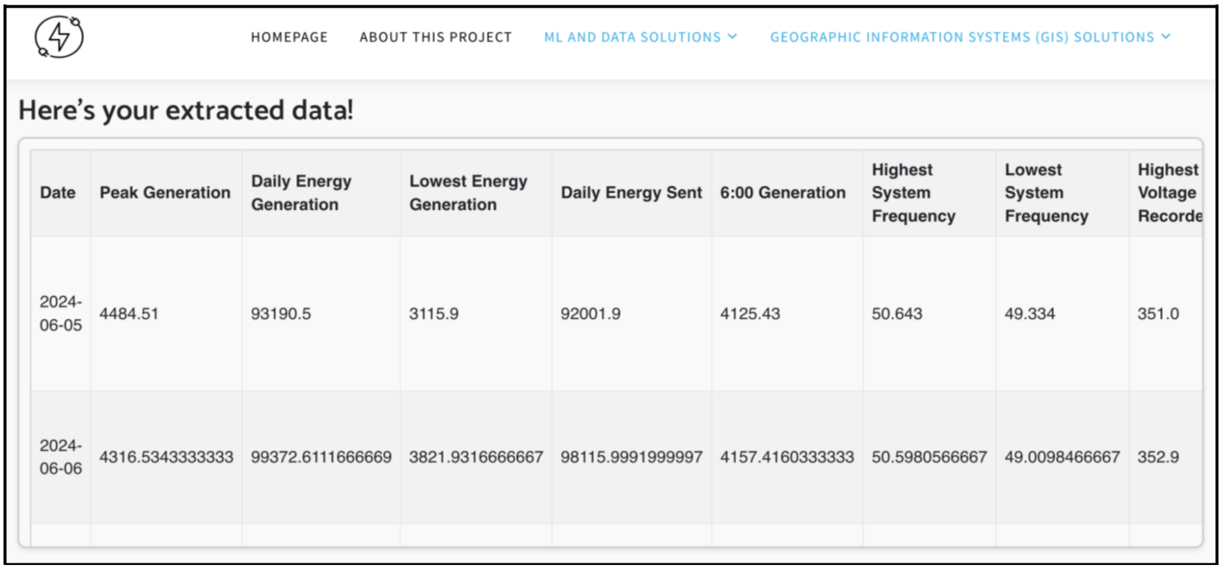


Figure 4.7: Selecting Start and End Dates Dynamically Adds Locations to Make Requests

Upon clicking "Fetch Data," the extracted information is presented, allowing users to download the CSV file containing both system operations and weather data.



Date	Peak Generation	Daily Energy Generation	Lowest Energy Generation	Daily Energy Sent	6:00 Generation	Highest System Frequency	Lowest System Frequency	Highest Voltage Recorde
2024-06-05	4484.51	93190.5	3115.9	92001.9	4125.43	50.643	49.334	351.0
2024-06-06	4316.5343333333	99372.6111666669	3821.9316666667	98115.9991999997	4157.4160333333	50.5980566667	49.0098466667	352.9

Figure 4.8: Retrieving Systems and Weather Data for Selected Timeframe and Locations

Outage Prediction Model

1. Model Performance and Evaluation

For predicting outages in the IKEDC Distribution Network, three models were evaluated: Logistic Regression, Random Forest, and Neural Network. Logistic Regression was initially chosen for its simplicity and interpretability. However, it showed lower performance metrics compared to the other models tested.

Random Forest demonstrated superior ability in accurately predicting unplanned outages, such as identifying a feeder line outage caused by a high-voltage fault, but also in effectively distinguishing non-outage instances. Neural Networks, known for their potential to capture intricate data patterns, were not pursued further due to the time-intensive process required for robust development within project deadlines. Developing a Neural Network to achieve comparable results to Random Forest would have required extensive fine-tuning and computational resources.
The results of the random forest model for both classes are given below:

Table 4.3: Performance Metrics and Results of the Outage Prediction Model

Metrics	Outage	Non-Outage	Overall
Precision	0.71	0.77	0.74
Recall	0.68	0.79	0.73
Accuracy or F1 Score	0.70	0.78	0.74

The evaluation was conducted using a dedicated test set, ensuring that the performance metrics reflect how well the models generalise to new, unseen instances of outages and non-outages data.

A model with a higher AUC score indicates better overall performance in distinguishing between outages and non-outages across all possible thresholds. This suggests that the model is effective in correctly identifying outages while minimising false alarms (predicting non-outages as outages).
The outage prediction model obtained an AUC score of 0.82, interpreting strong ability to distinguish between positive (outage) and negative (non-outage) instances.

After achieving impressive results in terms of precision, recall, F1 score, and AUC, the Mean Squared Error (MSE) was also evaluated on the test set to provide a comprehensive assessment of the model's performance. The Random Forest model for predicting power outages achieved a Mean Squared Error (MSE) of 0.257.
The MSE of 0.257 indicates that, on average, the squared difference between the model's predictions and the actual outcomes is 0.257. This relatively low MSE signifies that the model's predictions are close to the actual values. Combined with the overall precision of 0.74, recall of 0.73, F1 score of 0.74, and a high AUC of 0.82, the low MSE further validates the robustness and reliability of the model in predicting power outages.

Neural Networks, known for their potential to capture intricate data patterns, were not pursued further due to the time-intensive process required for robust development within project deadlines. Developing a Neural Network to achieve comparable results to Random Forest would have required extensive fine-tuning and computational resources.

2. Integrating Into The Web Application

The model was deployed onto a webpage in the web application with the **"/predict-outage"** endpoint.
The webpage is seen below:

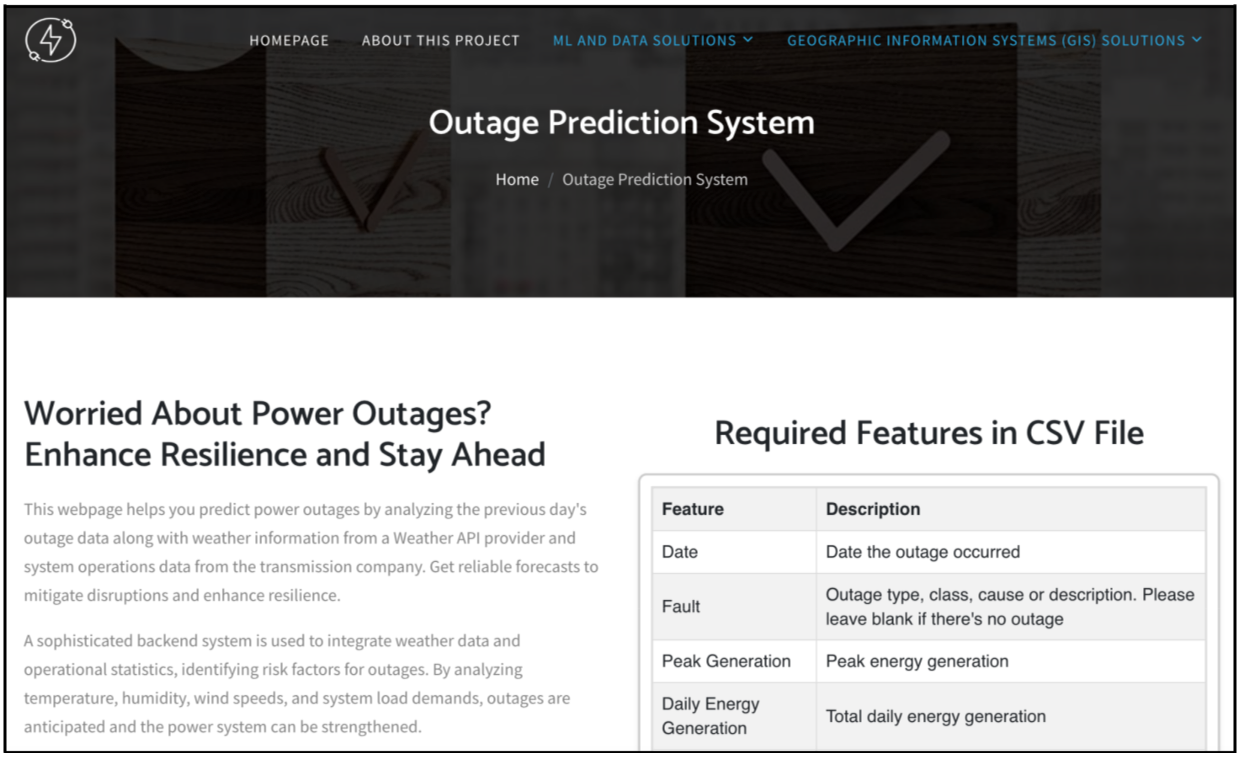


Figure 4.9: First Look at the Outage Prediction Model Web Application

To predict outages on the web application, a number of different features are required from the CSV file to be uploaded. Please see Table A1 for more information on these features.

A sample CSV file containing these column names is provided by the webpage. This process facilitates utilities in leveraging predictive analytics to anticipate and prepare for potential power outages more effectively. Next, the CSV file is uploaded containing the above fields, in this case **complete_data.csv**:

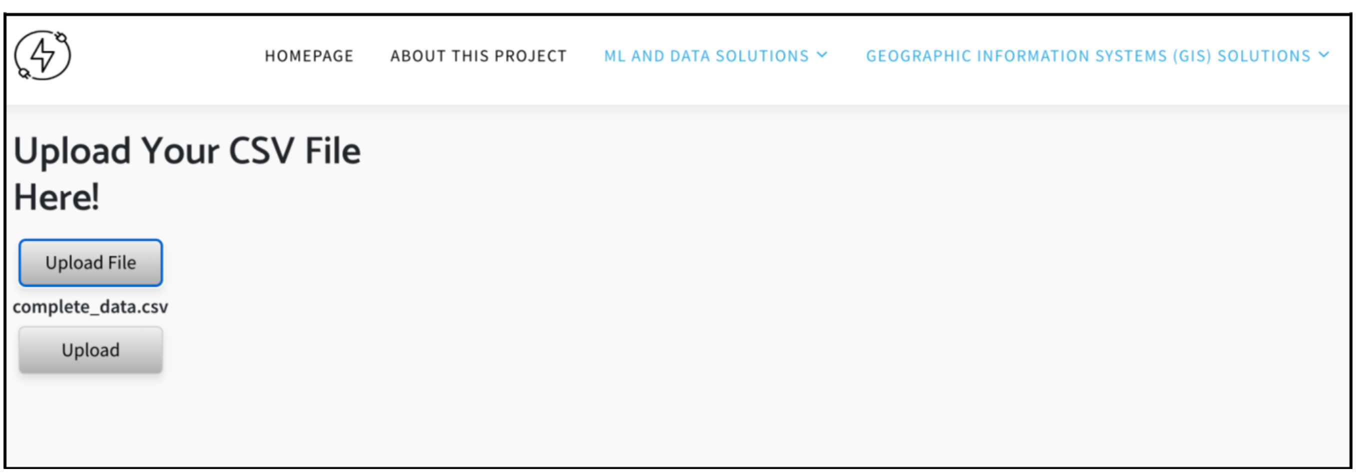


Figure 4.10: Testing the Upload Feature of the Outage Prediction System

After the file is uploaded, the outage prediction model runs and makes predictions upon the data to generate prediction results (predicting whether an outage will occur the next day and its probability in %), as seen below:

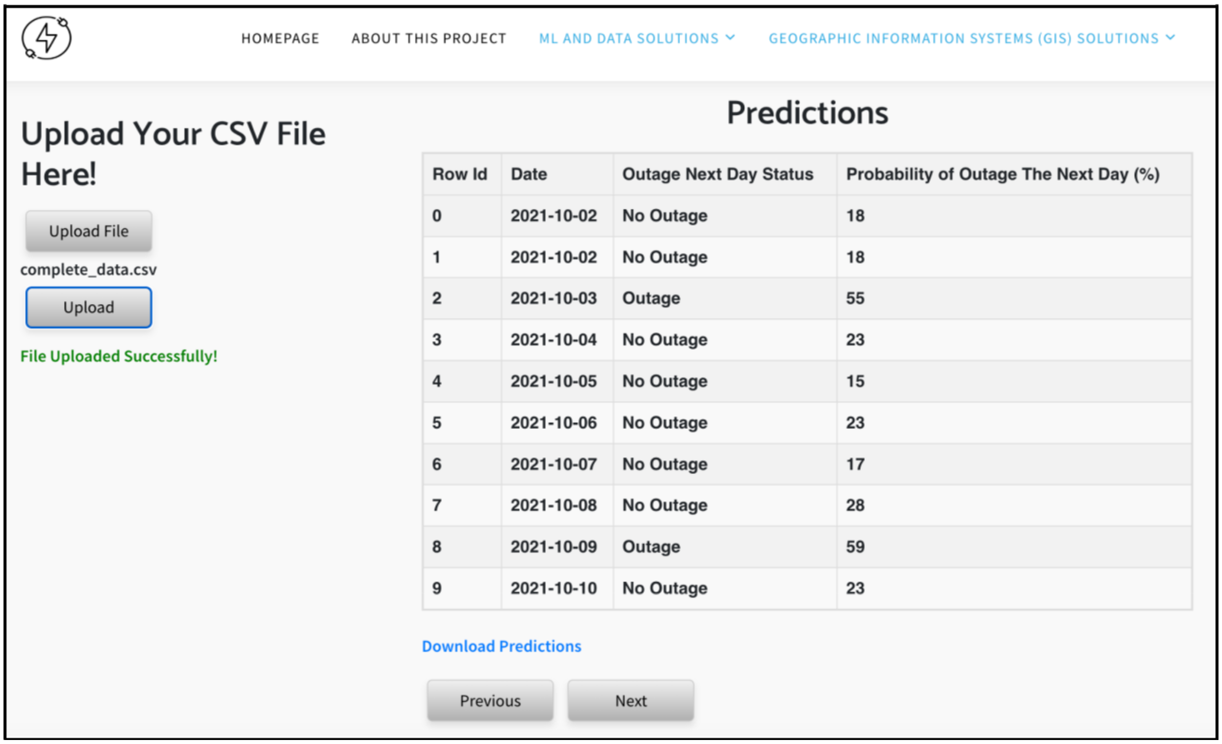


Figure 4.11: Testing the Outage Prediction Model with a Sample CSV File

After making predictions, there is the option of downloading the results as a CSV which can be used for further analysis, informing proactive maintenance strategies, and optimising resource allocation to mitigate potential power outages. This data-driven approach enhances operational resilience and helps in minimising service disruptions for customers.

Outage Classification Model

1. Model Performance and Evaluation

For classifying outage types in the IKEDC Distribution Network, various models were evaluated, focusing on accurately categorising fault descriptions into specific categories. The models considered include Logistic Regression, Random Forest, and Support Vector Machines (SVM). Random forest once again demonstrated superior ability in accurately classifying various fault descriptions into specific outage categories. The results for each category from the Random Forest model are summarised in Table 4.4:

Table 4.4: Performance Metrics and Results Across Different Outage Classes

Outage Class	**Precision**	**Recall**	**F1-Score**
Breaker	1.00	0.75	0.86
Customer Notification	0.94	0.89	0.92
Feeder	0.83	0.80	0.81
High-Voltage (Distribution) System	1.00	1.00	1.00
Human	0.83	0.83	0.83
Planned	0.57	0.89	0.70
Substation	1.00	1.00	1.00
Transformer	0.89	0.89	0.89
Transmission	1.00	0.60	0.75
Overall	0.87	0.85	0.85

The outage classification model's initial metrics - precision, recall, and F1 score show promising results, indicating good performance in predicting outage classes. However, given the class imbalance in the dataset, it's crucial to evaluate additional metrics for a deeper understanding of the model's performance.

To further explore the model's performance, the confusion matrix was examined. The confusion matrix reveals the distribution of correct and incorrect predictions across all classes as seen below. This detailed breakdown helps identify specific areas where the model performs well or struggles.

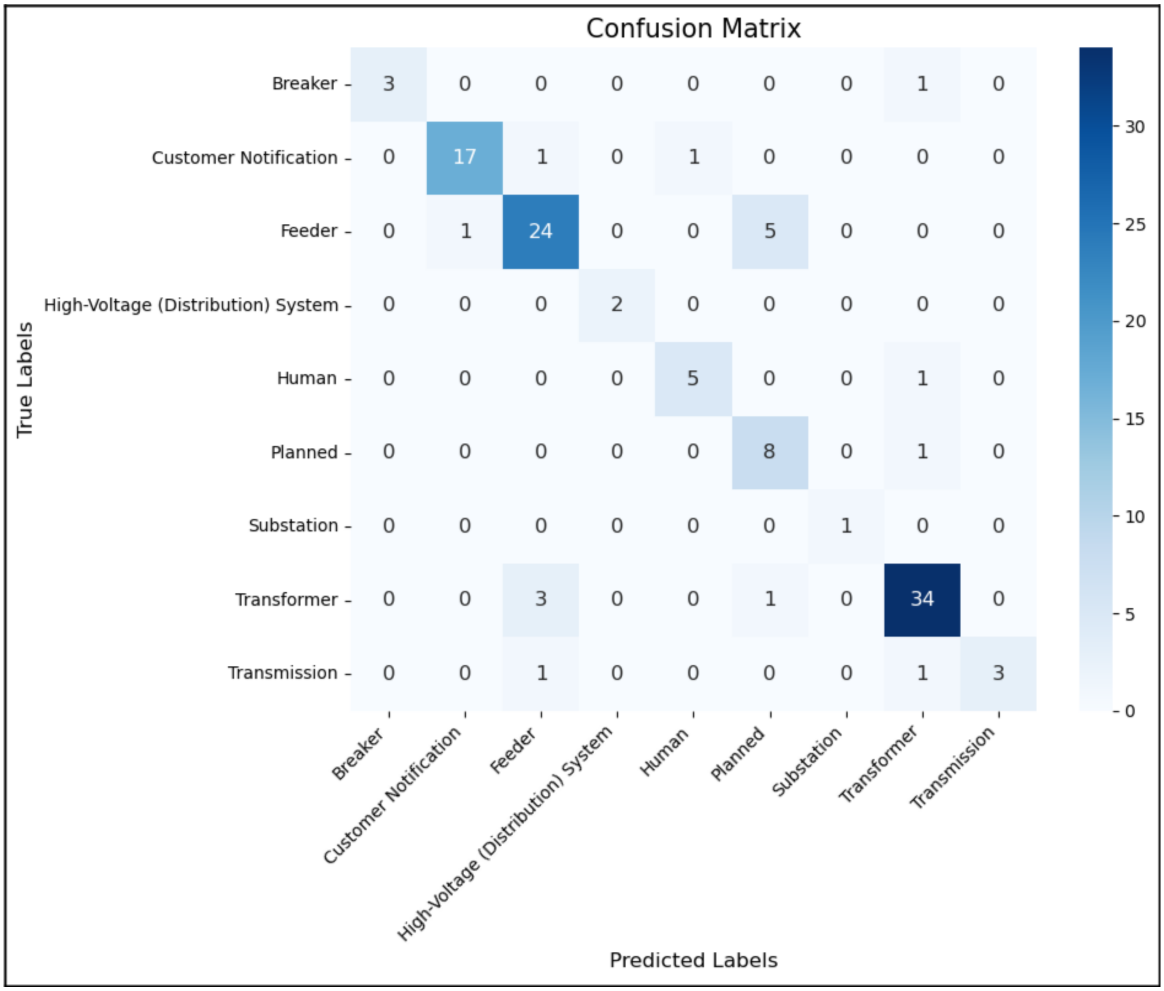


Figure 4.12: Confusion Matrix for the Outage Classification Model

In terms of agreement between the predicted and true labels, Cohen's Kappa score is 0.81. This score indicates good agreement, suggesting that the model's predictions are very reliable. Similarly, the Matthews Correlation Coefficient (MCC) is 0.81, which indicates a strong correlation between the predicted and actual classes, taking into account the size of the classes and their balance.

2. Integrating Into The Web Application

The model was deployed onto a webpage in the web application with the **"/classify-outage"** endpoint.
The webpage is seen below:

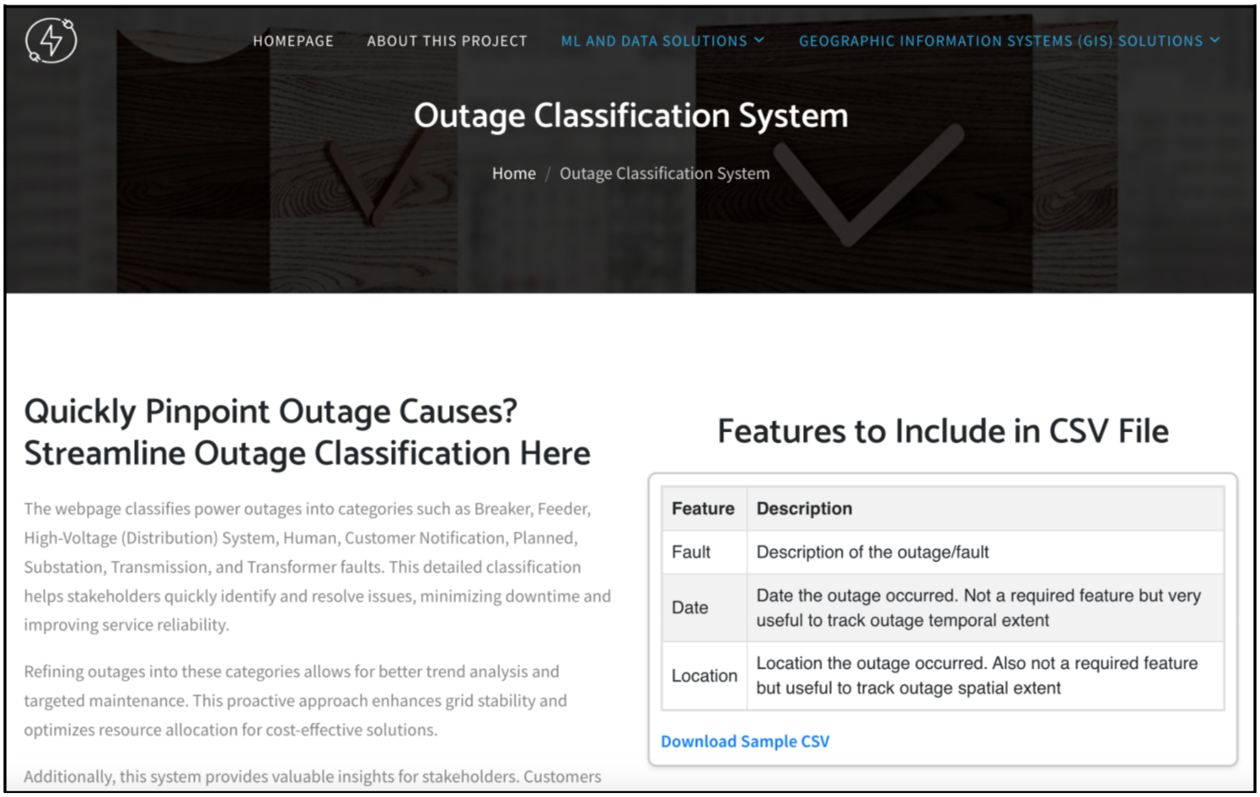


Figure 4.13: First Look at the Outage Classification Model Web Application

To classify outages on the web application, some features are used from the uploaded CSV file. Please refer to Table A2 for more information.

A sample CSV file containing these column names is also provided by the webpage for user convenience. Next, the CSV file containing the above fields is uploaded, named **complete_data.csv**, as in Figure 4.8. After the file is uploaded, the outage classification model runs and makes predictions upon the data to generate prediction results (outage type), as seen below:

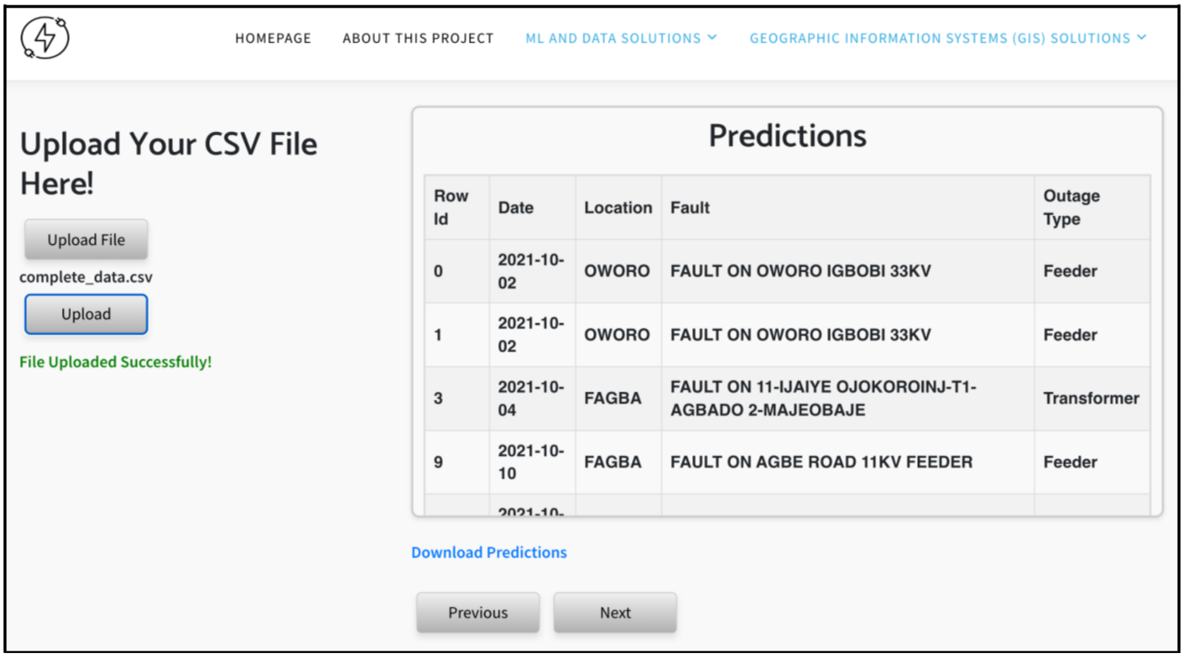


Figure 4.15: Testing the Outage Classification Model with a Sample CSV File

After making predictions, the web application allows users to download the results as a CSV file. This feature is useful for further analysis, enabling utilities and electrical engineers to review and classify faults by outage classes. It aids in identifying patterns, improving outage management, and developing more effective mitigation strategies.

Outage Mapping System (GIS Solutions)

Initial data exploration involved analysing outage logs and weather data to identify patterns and correlations. The data included timestamps, locations, and descriptions of outages.

Figure 4.15: First Look at the Outage Mapping System Web Application

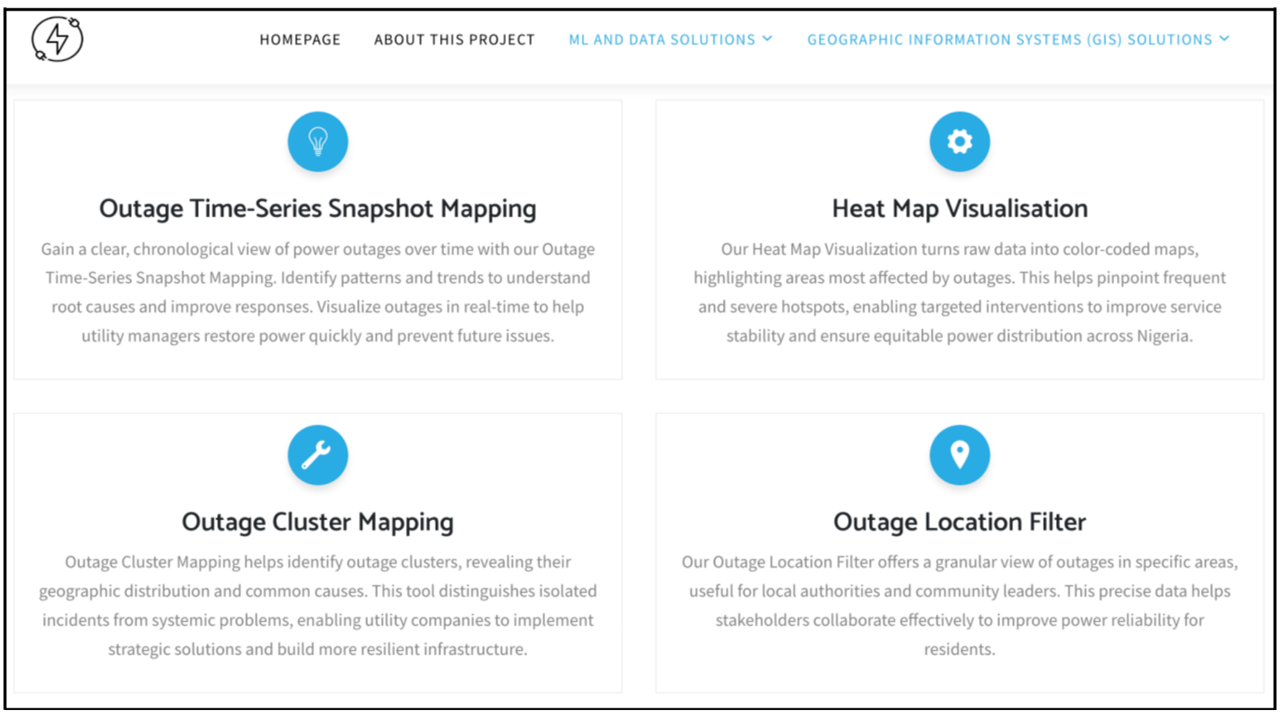


Figure 4.16: Scope of the Outage Mapping System Web Application

1. Heat Map Visualisation

Heat Map Visualisation transformed raw outage data into colour-coded maps, highlighting the most affected areas. This visualisation technique helps electrical engineers and utilities pinpoint frequent and severe outage hotspots, facilitating targeted interventions and infrastructure improvements.

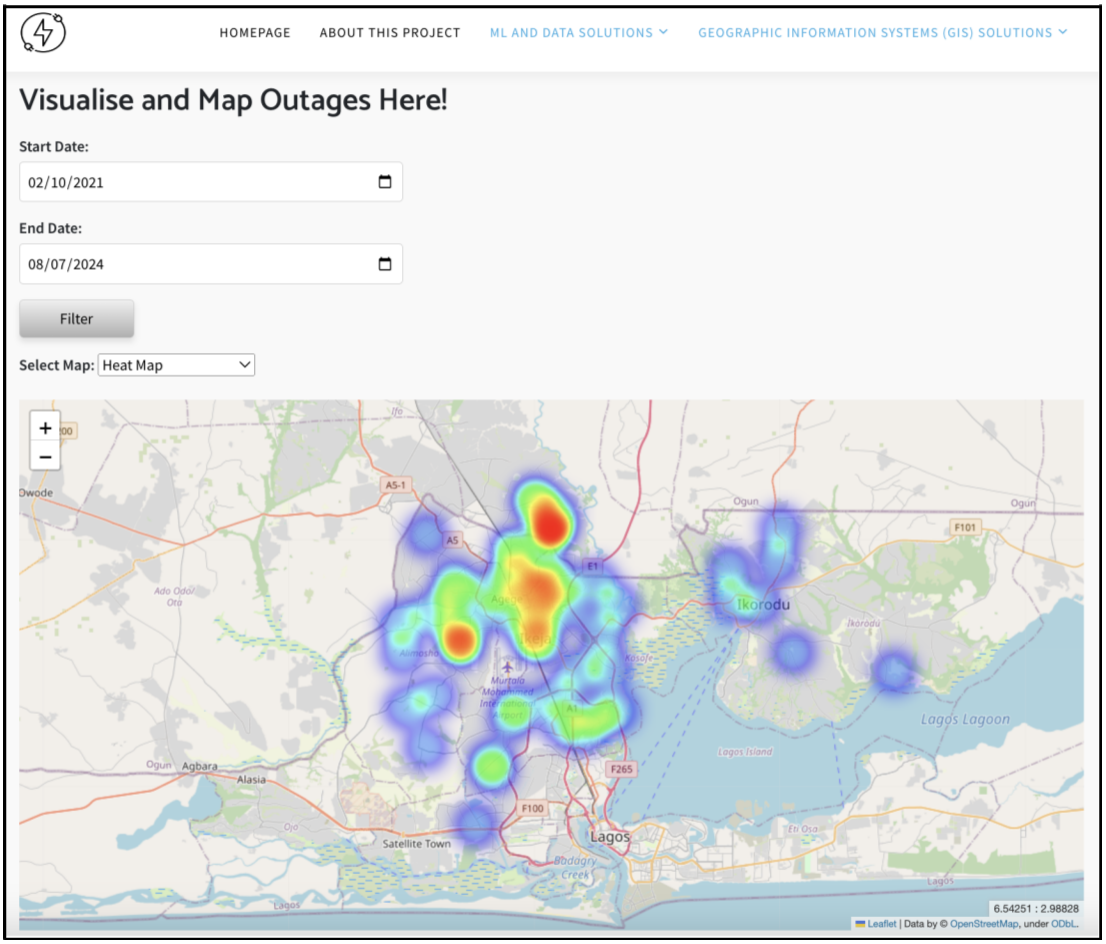


Figure 4.17: Holistic View of Lagos, Nigeria's Outage Heat Map on OpenStreetMap

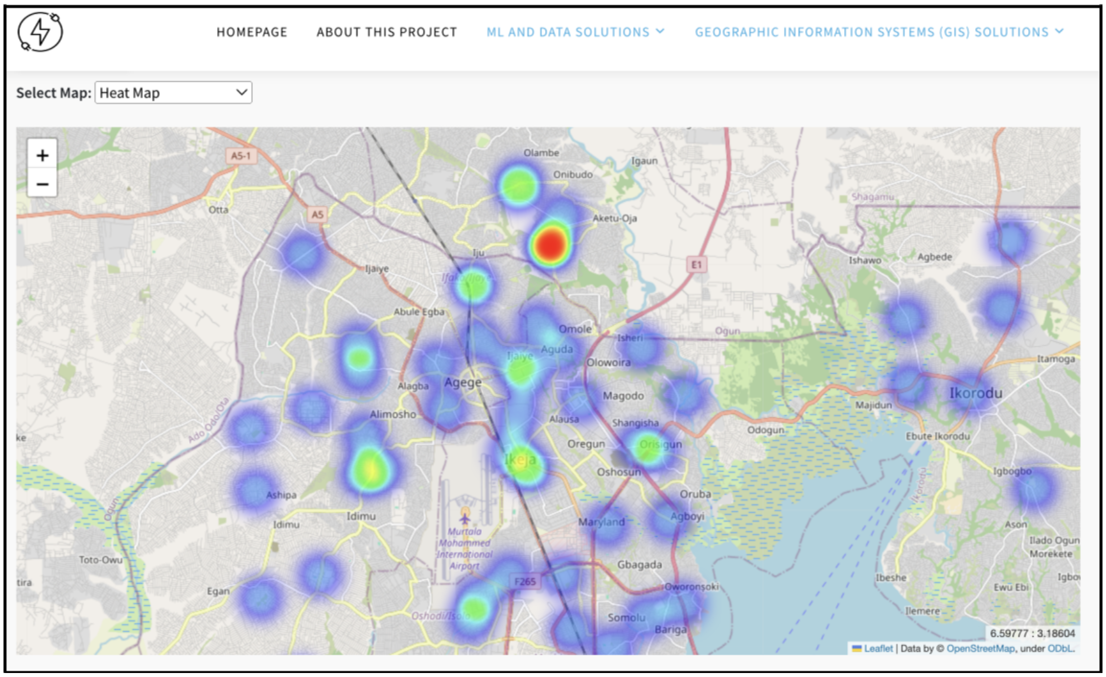


Figure 4.18: Outage Occurrences in Landmarks and Localities Using the Heat Map

The heat maps clearly indicated regions with the highest outage frequencies, helping utility managers and electrical engineers prioritise these areas for maintenance and upgrades. This visualisation also evaluated the equity of power distribution across the covered branches by identifying regions that required additional resources to stabilise the power supply. This evaluation was crucial for planning and optimising the performance of the electrical grid.

2. Outage Cluster Visualisation

Outage Cluster Mapping identified and analysed outage clusters, revealing their geographic distribution and common causes as seen below:

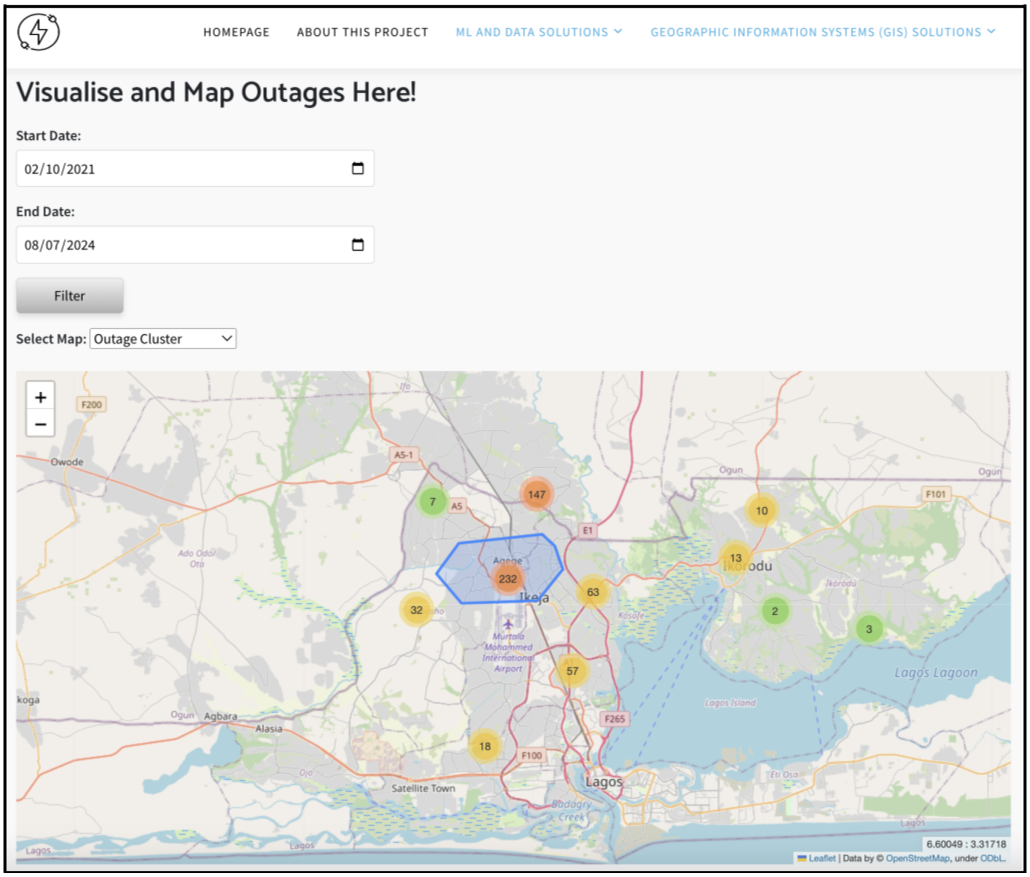


Figure 4.19: Cluster Map Showing Outage Distribution for Each Branch and Spatial Extent of a Sample Branch (Akowonjo) Represented As The Hexagonal Multipolygon

The colour-mapped number of outages can be observed, with green indicating relatively low outages, yellow indicating medium outages, and red indicating hotspot zones. Notably, Iju and Akowonjo are hotspot areas, while Ikorodu is relatively stable. Outages in Ikeja are moderate and not a significant concern.
The map also allows for interactive zooming in and out, enabling a more detailed analysis of outage numbers in specific local areas as seen below:

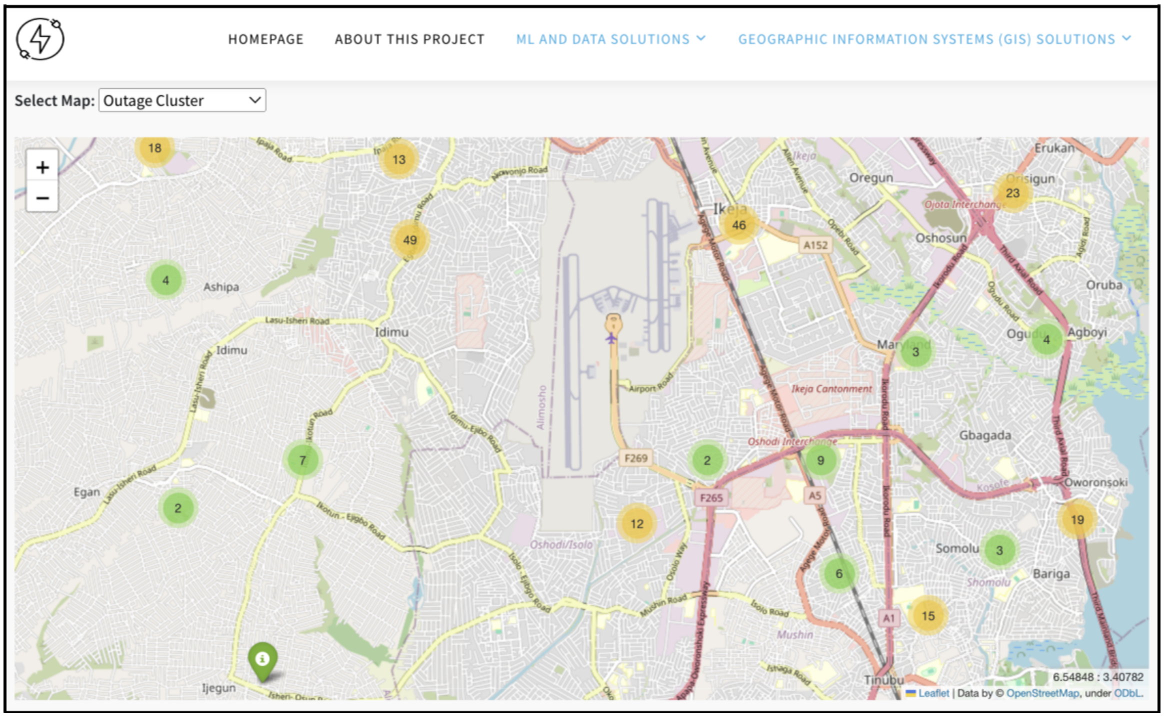


Figure 4.20: Interactive Cluster Map with Zoom Functionality Showing Color-Coded Outage Numbers

When outages are broken down to a specific undertaking, each instance can be detailed with the date and time of the outage, fault cause, and branch, as shown below:

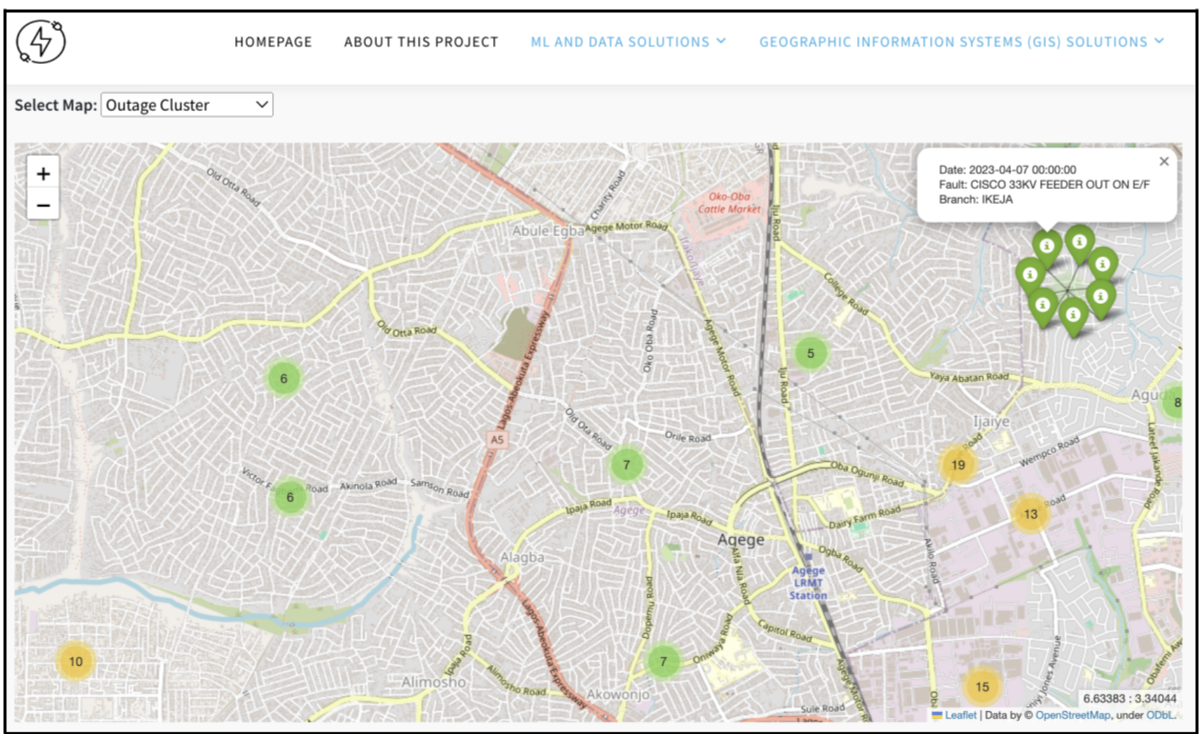


Figure 4.21: Detailed Breakdown of Outage Instances by Date, Time, Fault Cause, and Branch for a Specific Undertaking

This tool is essential for electrical engineers to distinguish between isolated incidents and systemic problems within the power distribution network. The clustering analysis showed that certain areas, particularly in Iju and Agege, consistently experienced higher outage rates, suggesting underlying infrastructure issues. Electrical engineers can use this information to focus on areas needing significant improvements.

3. Outage Location Filter

The Outage Location Filter provided a granular view of areas affected by outages, offering detailed insights for electrical engineers and local authorities.

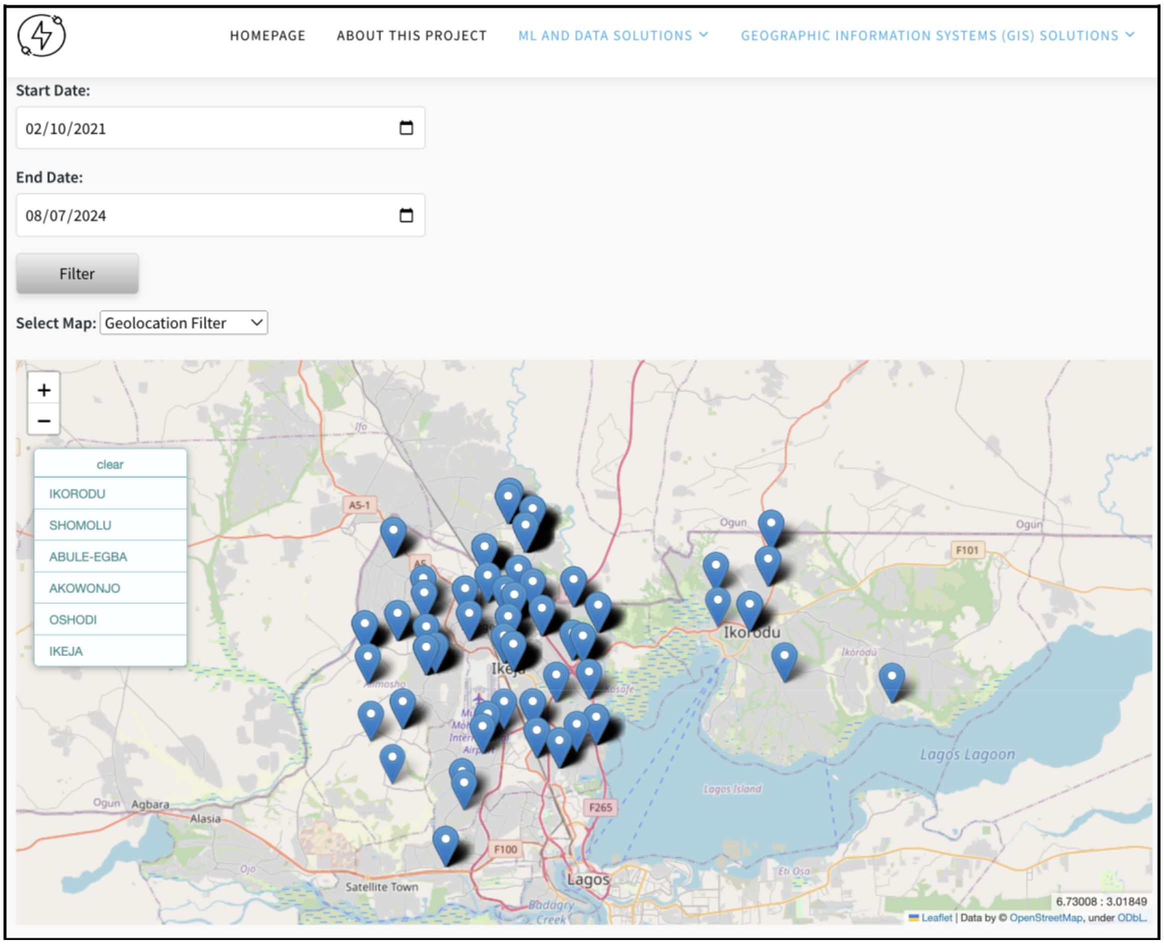


Figure 4.22: Geolocation Filter Side-Bar Before Filtering Outages by Location/Branches

Next, the geofilter web-system is tested by using the side-bar filter for a sample location (Ikorodu in this case). The filter tool

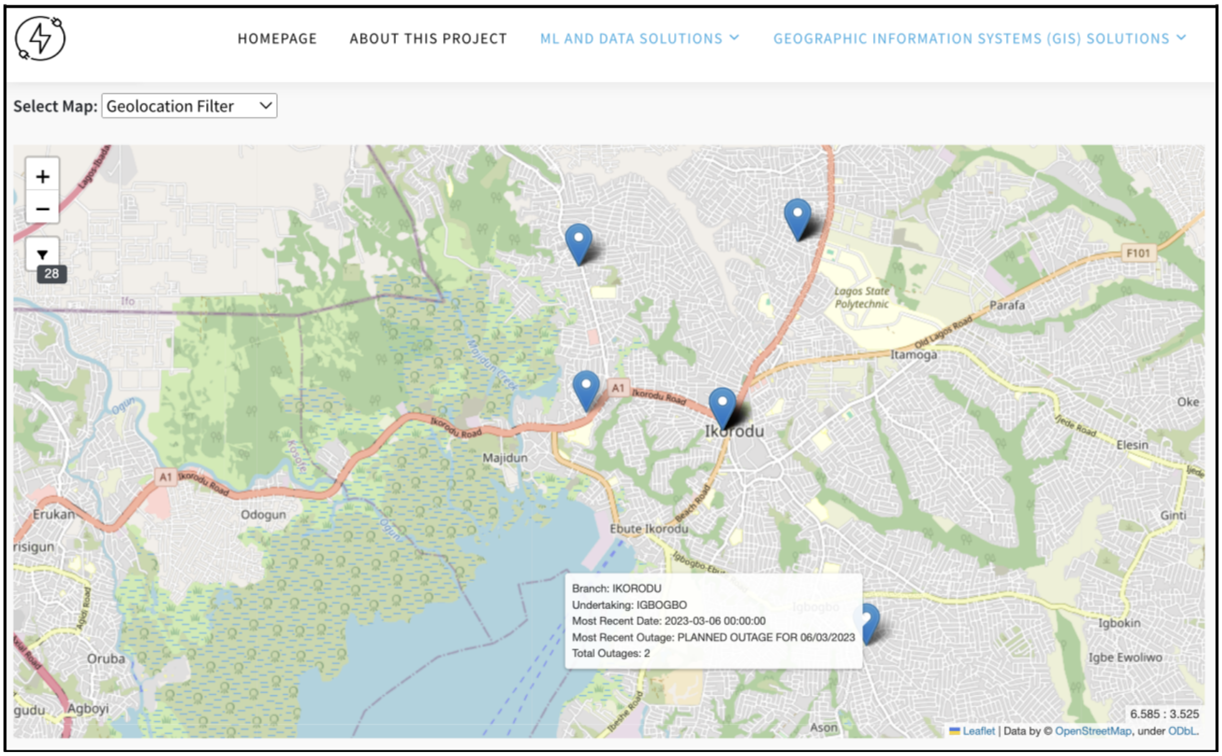


Figure 4.23: Geolocation Filter Side-Bar After Filtering Outages by a Branch (Ikorodu)

The filter tooltip displays a total of 28 outages for Ikorodu as of July 9th, 2024. When exploring an undertaking, users can view details such as the date of the most recent outage, its cause, and the total number of outages within the specified time period as seen above.

The location filter allowed for the examination of outages at a highly detailed level, down to specific branches. This precision was essential for targeted maintenance and rapid response to outages. This detailed geoinformatic map facilitates collaboration among stakeholders, including utility companies and local governments, to improve power reliability for residents. Electrical engineers could leverage this data to design and implement more effective outage mitigation strategies. Common causes of outages within clusters were identified, such as equipment failures, feeder issues, allowing for strategic interventions.

4. Outage Time-Series Snapshot Mapping

The Outage Time-Series Snapshot Mapping component offered a chronological view of power outages, enabling the identification of patterns and trends over time. This visualisation tool helped to understand the root causes of outages and improve response efforts. The timestamped map provides a visual narrative, presenting a video chronology of outage occurrences. It tracks the number of outages for each undertaking and records the most recent outage during the specified period, as illustrated below:

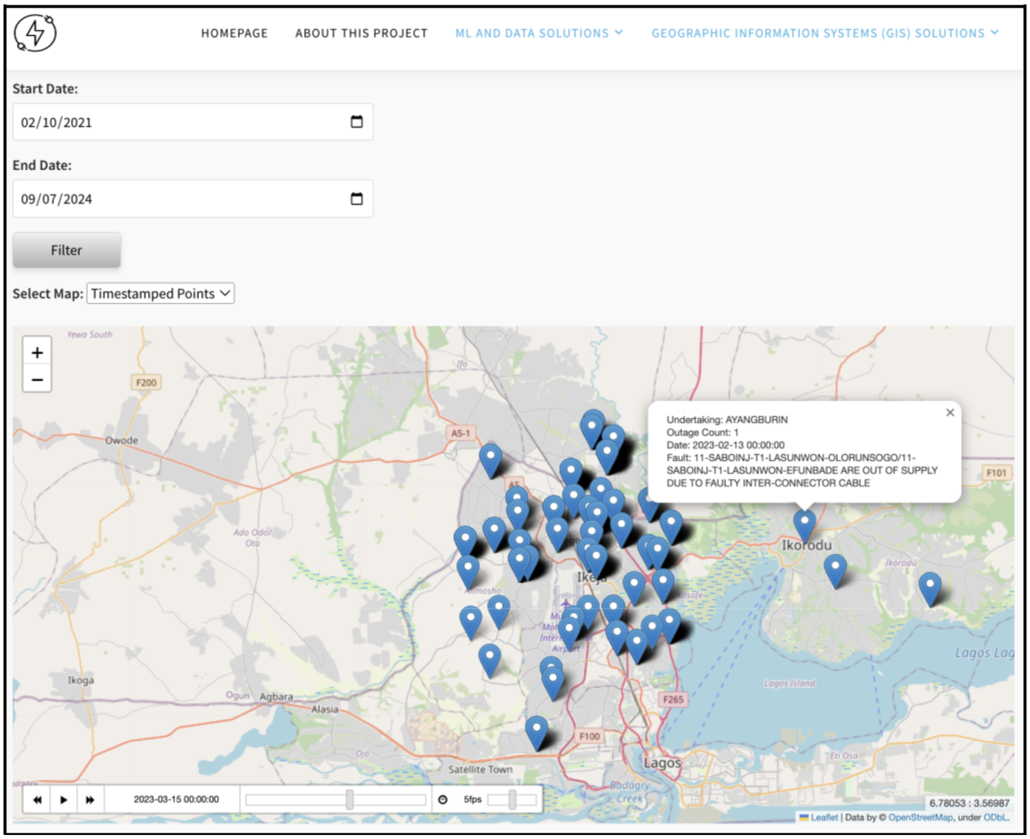


Figure 4.24: Timestamped Map Showing the Distribution of Outages with Time, Just Like a Movie (As of 15th March, 2023)

The map includes a pause and play feature located at the lower left, which are self-explanatory. Users can also adjust the speed or Frames Per Second (FPS), set to 5fps in the test example, to make the video faster or slower depending on the desired pace of analysis.

As the video progresses and undertakings record their initial outages, markers increase. The outage count for undertakings with existing outages rises based on the total number of outages to date. For example, as shown in Figure 4.25, the outage count for Ayangburin was 1 as of 15th March, 2023, and increased to 9 by 26th January, 2024. The most recent outage at Ayangburin was also updated accordingly.

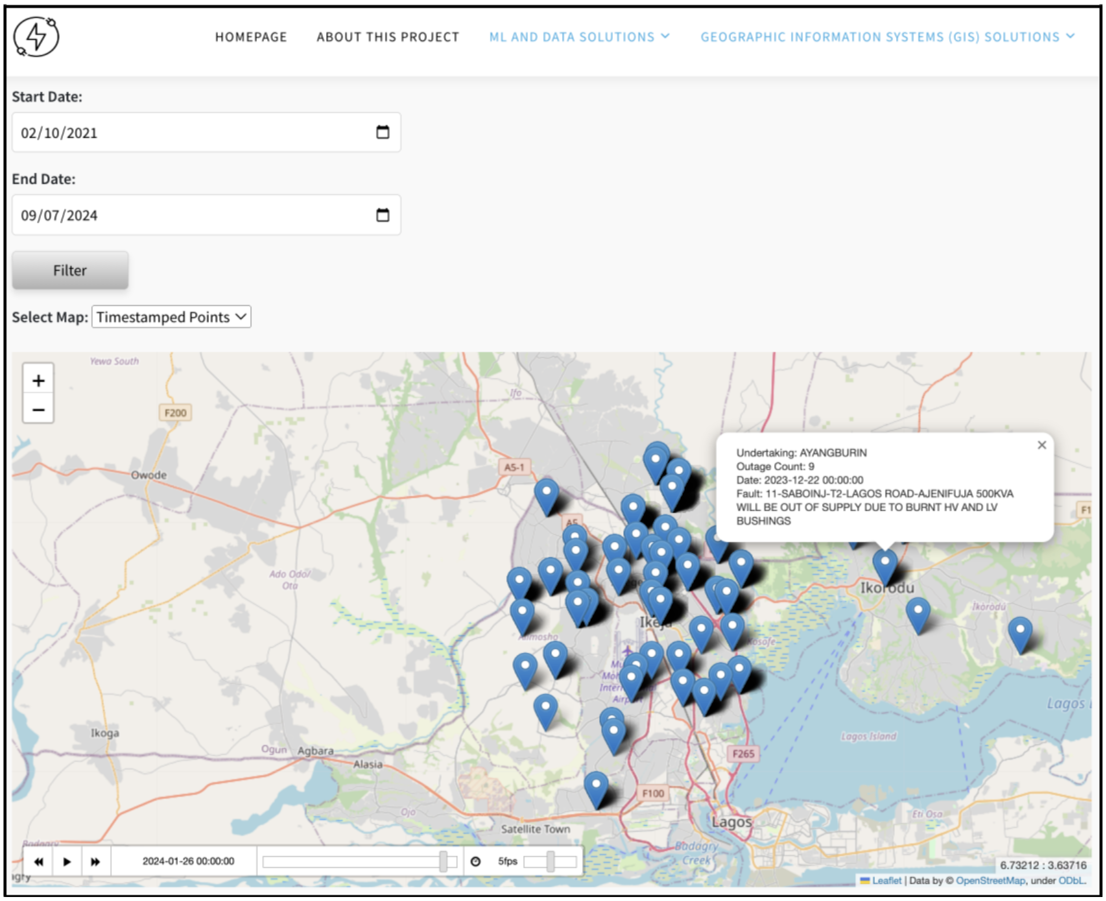


Figure 4.25: Timestamped Map Showing the Distribution of Outages with Time (As of 26th January, 2024)

This functionality helps electrical engineers to better understand outage patterns, improve response strategies, and enhance overall grid resilience. This trend analysis was crucial for proactive planning and resource allocation.

By examining the chronological data, several root causes of frequent outages were identified, including infrastructure weaknesses and extreme weather events.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

Summary of Findings

The project aimed to enhance the resilience of the distribution network by leveraging ML and GIS techniques to identify unplanned High Impact Low Probability (HILP) outages. Several systems were developed and validated, including an outage prediction system, outage classification system, integration of system operations and weather data, and an outage mapping system.

The analysis revealed that the majority of outages were unplanned, with 'Transformer' and 'Feeder' faults being the most common. It was observed that extreme weather conditions, specifically temperature extremes and low sea level pressure, significantly correlated with the frequency of outages. The Systems Operations data analysis indicated high transmission efficiency, suggesting that distribution side issues are the primary cause of outages.

The Random Forest model demonstrated superior predictive performance compared to Logistic Regression and initial Neural Network model, achieving an overall precision of 0.74, recall of 0.73, F1 score of 0.74, and an AUC score of 0.82. For outage classification, the Random Forest model effectively categorised fault descriptions with an overall precision of 0.87, recall of 0.85, and F1 score of 0.85.

The web application developed for outage prediction and classification allowed for user-friendly interaction with the models, facilitating proactive maintenance strategies and optimised resource allocation. The outage mapping system, featuring heat maps, cluster visualisations, location filters, and time-series snapshots, provided valuable insights into outage patterns and trends, enabling targeted interventions and infrastructure improvements.

Conclusion

The implementation of GIS techniques and machine learning models in this project has significantly improved the ability to predict and classify outages in the IKEDC Distribution Network. The integration of system operations and weather data has further enhanced the accuracy and reliability of these models. The visualisation tools developed provide a comprehensive understanding of outage patterns, facilitating effective decision-making and strategic planning.

The findings underscore the importance of focusing on distribution side issues to enhance power resilience. The strong correlation between weather conditions and outages highlights the need for resilient infrastructure capable of withstanding extreme weather events. The success of the Random Forest model in both prediction and classification tasks demonstrates the potential of machine learning in improving power distribution resilience.

Recommendations

To further enhance the resilience of electrical power systems, it is recommended to prioritise the resolution of distribution side faults, particularly those related to transformers and feeders. Regular maintenance and upgrades of these components should be conducted to minimise the likelihood of outages.

Given the significant impact of weather conditions on outage frequency, it is crucial to invest in weather-resistant infrastructure and adopt proactive measures to mitigate the effects of extreme weather. This includes reinforcing power lines, improving the durability of transformers, and implementing advanced monitoring systems to detect and respond to weather-induced faults promptly.

The web application developed in this project should be modified to suit the daily operations of other distribution networks, enabling continuous monitoring and prediction of outages. Training for utility managers and electrical engineers on the use of this application will ensure its effective utilisation.

Further Studies

Future research should explore the development of more sophisticated machine learning models, such as deep learning techniques, to further improve the accuracy and reliability of outage predictions and classifications. Additionally, incorporating real-time data from smart grid technologies could enhance the responsiveness of the models and enable dynamic adjustments based on current network conditions.

Studies should also investigate the long-term impacts of climate change on power distribution networks, aiming to develop adaptive strategies that can cope with evolving weather patterns. Collaborating with meteorological agencies to obtain high-resolution weather forecasts could further refine the predictive capabilities of the models.

Finally, expanding the scope of the outage mapping system to include customer feedback and social media data could provide a more comprehensive view of outage impacts, enabling more targeted and effective interventions. This holistic approach will contribute to a more resilient and reliable power distribution network, ultimately improving service delivery and customer satisfaction.

REFERENCES

[1] M. V. Eeeten, A. Nieuwenhuijs, E. Luiijf, M. Klaver, and E. Cruz, "The state and the threat of cascading failure across critical infrastructures: The implications of empirical evidence from media incident reports," *Public Adm.*, vol. 89, pp. 381–400, 2011.

[2] K. M. Kosa, S. C. Cates, S. L. Godwin, R. J. Coppings, and L. Speller-Henderson, "Most Americans are Not Prepared to Ensure Food Safety during Power Outages and Other Emergencies," *Food Prot. Trends*, vol. 31, pp. 428–436, 2011.

[3] J. O. Wear, "Is Your Hospital Ready for a Natural or Man-Made Disaster?" *IFMBE Proc.*, vol. 37, pp. 699–702, 2011.

[4] A. E. Airoboman, P. A. Amaize, A. E. Ibhaze, and O. O. Ayo, "Economic Implication of Power Outage in Nigeria: An Industrial Review," *International Journal of Applied Engineering Research*, vol. 11, no. 7, pp. 4930-4933, 2016.

[5] Amadi, H.N. (2015). "Impact of power outages on developing countries: evidence from rural households in Niger Delta, Nigeria." *Journal of Energy Technologies and Policy*, 5(3), 27-38.

[6] S. Akanni, Samson, A. Melodi, and T. Ale, "Analysis of the Causes of Outages in the Distribution System of Lagos State, Nigeria," *FUOYE Journal of Engineering and Technology*, vol. 8, 2023.

[7] J. I. Igbokwe and E. J. Emengini, "GIS in Management of Electricity Distribution Network: A case study of Onitsha-North L.G.A., Anambra state, Nigeria," *Department of Surveying and Geoinformatics, Faculty of Environmental Sciences, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria*, 2005. [Online]. Available: http://www.gisdevelopment.net/application/urban/overview/power/index.html.

[8] A. Alshawish and H. de Meer, "Risk mitigation in electric power systems: Where to start?" *Energy Inform.*, vol. 2, p. 34, 2019.

[9] Jesuovie, O., M. Edafe, and B.O. Onoriode, "Power supply and national development, 1980- 2012: The Nigeria experience," *International Journal of Humanities and Social Science*, vol. 4, no. 8, pp. 144-154, 2014.

[10] Global Network on Energy for Sustainable Development (GNESD), "Energy access-Making power sector reform work for the poor," *Johannesburg*, pp. 1-16, 2002.

[11] A. B. Onakoya, A. O. Onakoya, O. A. Jimi-Salami, and B. O. Odedairo, "Energy consumption and Nigerian economic growth: An empirical analysis," *European Scientific Journal*, vol. 9, no. 4, pp. 25-40, 2013.

[12] M. M. Ologundudu, "The impact of electricity supply on industrial and economic performance in Nigeria," *International Global Journal of Management and Business Research*, vol. 2, no. 4, pp. 9-28, 2015.

[13] E. Lionel, "The dynamic analysis of electricity supply and economic development: Lessons from Nigeria," *Journal of Sustainable Society*, vol. 2, no. 1, pp. 1-11, 2013.

[14] E. George and J. Oseni, "The relationship between electricity power and unemployment rates in Nigeria," *Australian Journal of Business and Management Research*, vol. 2, no. 2, pp. 10-19, 2012.

[15] D. Farquharson, Devynne, P. Jaramillo, and C. Samaras, "Sustainability implications of electricity outages in sub-Saharan Africa," *Nature Sustainability,* vol. 1, 2018.

[16] J. Temitope, L. Oderinde, and B. Ajayi, "Electricity Outages and Micro-Enterprise Performance in Selected Semi-Urban Areas in Lagos and Ogun States, Nigeria," *İktisat Fakültesi Mecmuası / Journal of the Faculty of Economics,* vol. 6, pp. 1-20, 2021.

[17] I. Samuel, J. Katende, J. Daramola, and S. A. Awelewa, "Review of System Collapse Incidences on the 330-kV Nigerian National Grid," *International Journal of Engineering Science Invention*, vol. 3, no. 4, pp. 55-59, 2014.

[18] O. S. Maliki, A. O. Agbo, A. O. Maliki, L. Ibeh, and C. O. Agwu, "Comparison of Regression Model and Artificial Neural Network Model for the prediction of Electrical Power generated in Nigeria," *Advances in Applied Science Research*, vol. 2, 2011.

[19] A. Saidi, Alia, N. Ramli, N. Muhammad, and L. Awalin, "Power outage prediction by using logistic regression and decision tree," *Journal of Physics: Conference Series*, vol. 1988, 2021.

[20] W. O. Taylor, D. Cerrai, D. Wanik, M. Koukoula, and E. N. Anagnostou, "Community power outage prediction modelling for the Eastern United States," *Energy Reports*, vol. 10, 2023.

[21] D. Cerrai, M. Koukoula, P. Watson, and E. N. Anagnostou, "Outage prediction models for snow and ice storms," *Sustainable Energy, Grids and Networks*, vol. 21, 2020.

[22] H. Hou, Y. Zhu, H. Geng, M. Li, Y. Xie, L. Zhu, and Y. Huang, "Spatial distribution assessment of power outage under typhoon disasters," *International Journal of Electrical Power & Energy Systems*, vol. 132, 2021.

[23] D. Wanik, E. Anagnostou, M. Astitha, B. Hartman, G. Lackmann, J. Yang, D. Cerrai, J. He, and M. Frediani, "A Case Study on Power Outage Impacts from Future Hurricane Sandy Scenarios," *Journal of Applied Meteorology and Climatology*, vol. 57, 2017.

[24] M. Kezunovic, Z. Obradovic, T. Djokic, and S. Roychoudhury, "Systematic Framework for Integration of Weather Data into Prediction Models for the Electric Grid Outage and Asset Management Applications," *Hawaii International Conference on System Sciences*, 2018.

[25] J. Cai, Y. Cai, H. Cai, S. Shi, Y. Lin, and M. Xie, "Feeder Fault Warning of Distribution Network Based on XGBoost," *Journal of Physics: Conference Series*, vol. 1639, 2020.

[26] A. R., Khosa, I., Armghan, A., Arshad, J., Rabbani, S., Alsharabi, N., and Hamam, H., "Financial Hazard Prediction Due to Power Outages Associated with Severe Weather-Related Natural Disaster Categories," *Energies*, vol. 15, no. 24, 2022.

[27] V. Sultan and B. Hilton, "A Spatial Analytics Framework to Investigate Electric Power-Failure Events and Their Causes," *ISPRS International Journal of Geo-Information*, vol. 9, 2020.

[28] S. Kumar, K. N., B. Kannan, K. G. Harish, and N. Jaya Anand, "A Novel Approach for the Three Phase Power Failure Prediction Model Using AI/ML," *International Journal of Innovative Science and Research Technology (IJISRT)*, vol. 8, no. 7, 2023.

[29] V. O. Eniola, E. A. Akeju, and O. F. Florence, "Mitigating poor electric distribution network in Nigeria via geographical information system application," *European Journal of Social Sciences Studies*, vol. 0, pp. 1-9, 2017. [Online]. Available: https://doi.org/10.5281/zenodo.123456.

[30] K. Hafeez and M. Khan, "Application of Geographic Information System for the Installation of Surge Arresters on overhead 132kv Power Line," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 2, 2012.

[31] Y. Kakumoto, Y. Koyamatsu, A. Shiota, Y. Qudaih, and Y. Mitani, "Application of Geographic Information System to Power Distribution System Analysis," *Energy Procedia*, vol. 100, 2016.

[32] A. Ünverdi and A. Sisman, "Developing an Algorithm on the Reporting Of Outages in the Electricity Distribution System with GIS Integration." *Intercontinental Geoinformation Days (IGD*, vol.1, 2020.

[33] W. Wu, P. Li, B. Wang, Y. Liu, T. Xu, H. Du, and Y. He, "Integrated Distribution Management System: Architecture, Functions, and Application in China," *J. Mod. Power Syst. Clean Energy*, vol. 10, pp. 245–258, 2022.

[34] Ethman, I., Maaroufi, Mohamed, Mahmoud, Abdel, YAHFDHOU, Ahmed, & Youm, Issakha. (2017). "Load Flow Study and Reactive Power Optimization of Nouakchott Power Grid." *IPASJ International Journal of Electrical Engineering (IIJEE)*, 5, 9-25.

[35] A. Farooq, A. Saad, and M. Zuhaib, "Midpoint Voltage Regulation of Transmission Lines and Compensation in Pakistan Scenario," *International Journal of Technology and Research*, vol. 4, pp. 21-24, 2015.

[36] G. Machado, M. Recamonde-Mendoza, and L. Corbellini, "What variables are important in predicting bovine viral diarrhoea virus? A random forest approach," *Veterinary Research*, vol. 46, p. 85, 2015. doi: 10.1186/s13567-015-0219-7.

[37] L. Anselin, "Local Indicators of Spatial Association - LISA," *Geographical Analysis*, vol. 27, pp. 93-115, 1995. doi: 10.1111/j.1538-4632.1995.tb00338.x.

[38] J. Zhang and Y. Luo, "Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network," *2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017)*, pp. 300-303, Mar. 2017. doi: 10.2991/msam-17.2017.68.

[39] M. Imrane, C. Chemseddine, N. Said, E. M. Bellfkih, M. S. el Kasmi Alaoui, and M. Abdelaziz, "Machine learning for decoding linear block codes: case of multi-class logistic regression model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 1, pp. 538-547, 2021. doi: 10.11591/ijeecs.v24.i1.pp538-547.

[40] H. Vo, S. Taremsari, B. Ranjbar, H. Mashhadimoslem, E. Rahimi, M. Rahimi, and A. Elkamel, "Hydrogen Storage on Porous Carbon Adsorbents: Rediscovery by Nature-Derived Algorithms in Random Forest Machine Learning Model," *Energies*, vol. 16, 2023, doi: 10.3390/en16052348.

[41] J. Thomas, T. Maszczyk, N. Sinha, T. Kluge, and J. Dauwels, "Deep learning-based classification for brain-computer interfaces," *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 234–239, 2017.

APPENDIX

Features Required for the Outage Prediction and Classification Models

This appendix describes the different features required from the CSV file to be uploaded for outage prediction and classification. These were the features the model performed best with after rigorous testing and feature extraction:

Table A1: Descriptions of Required Features for the Outage Prediction Model

Feature	**Description**
Date	Date the outage occurred
Fault	Outage type, class, cause, or description. Please leave blank if there's no outage
Peak Generation	Peak energy generation
Daily Energy Generation	Total daily energy generation
Lowest Energy Generation	Lowest energy generation
Daily Energy Sent	Total daily energy sent
6:00 Generation	Energy generation at 6:00
Highest System Frequency	Highest system frequency
Lowest System Frequency	Lowest system frequency
Highest Voltage Recorded	Highest voltage recorded
Lowest Voltage Recorded	Lowest voltage recorded
temp	Temperature at the location (°C). Daily values are average values (mean) for the day.
feelslike	What the temperature feels like accounting for heat index or wind chill (°C). Daily values are average values (mean) for the day
humidity	Relative humidity in %
precip	The amount of liquid precipitation that fell or is predicted to fall in the period (millimetres)
precipprob	The likelihood of measurable precipitation ranging from 0% to 100%
precipcover	The proportion of hours where there was non-zero precipitation (%)

windspeed	The instantaneous wind speed at a location (Kilometres per hour). May be empty if it is not significantly higher than the wind speed. Daily values are the maximum hourly value for the day
winddir	Direction from which the wind is blowing (degrees from North)
sealevelpressure	The sea level atmospheric or barometric pressure in millibars (or hectopascals)
visibility	Distance that can be seen in daylight (kilometres)
solarradiation	(W/m2) the solar radiation power at the instantaneous moment of the observation (or forecast prediction)
solarenergy	(MJ /m2) indicates the total energy from the sun that builds up over an hour or day
uvindex	A value between 0 and 10 indicating the level of ultraviolet (UV) exposure for that hour or day. 10 represents high level of exposure, and 0 represents none

Table A2: Descriptions of Features for the Outage Classification Model

Feature	**Description**
Fault	Description of the outage/fault for classification
Date	Date the outage occurred. Not a required feature but very useful to track outage temporal extent
Location	Location the outage occurred. Also not a required feature but useful to track outage spatial extent