BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

NAME Trish

SURNAME. Mapiku

REG NUMBER B201177B

COURSE CODE IT 414

PROGRAM NAME

BACHELOR OF SCIENCE HONOURS DEGREE IN INFORMATION TECHNOLOGY

RESEARCH PROJECT TITLE

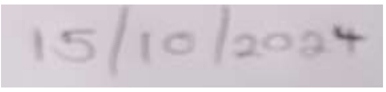
Application of artificial intelligence (AI) and machine learning for proactive network troubleshooting and fault prediction (case study ZETDC HARARE)

SIGNED APPROVAL FORM

The undersigned certify that they have read and recommended to the Bindura University of Science Education for the acceptance of a dissertation entitled **"Application of artificial intelligence (AI) and machine learning for proactive network troubleshooting and fault prediction (case study ZETDC HARARE).**

……………………………………… ……………………………………

Signature of student **date**

……………………………………… 

……………………………………

Signature of the supervisor **date**

……………………………………… 

……………………………………

Signature of department chairperson **date**

DEDICATION

My grandmother Viola Makumire, who served as my motivation and constantly prayed for me, is remembered in my dissertation. Additionally, I dedicate this research to my family, who have always supported me without fail, and to my close friends, who have always been there for me when I needed them most. Finally, I dedicate this dissertation to the All-Powerful God who created everything and brought me to this point in my life. With you, nothing is insurmountable. I praise the Lord.

ACKNOWLEDGEMENTS

First and foremost, I would want to thank my parents, my brother Trevor Mapiku and my friends for their financial and emotional support respectively. Secondly, I thank my supervisor, Mr Musariwa, for his support, both material and emotional, during this study endeavour. Finally, I thank God for providing me with the bravery and fortitude I needed to finish my studies.

ABSTRACT

Zimbabwe Electricity Transmission and Distribution Company (ZETDC) is rooted in the dynamic landscape of the electricity industry in Zimbabwe. As the primary electricity distribution and transmission utility in the country, ZETDC faces a confluence of challenges and opportunities. This study is geared toward exploring strategies and solutions that will bolster the reliability, efficiency, and sustainability of electricity distribution and transmission, ensuring that ZETDC continues to play a vital role in Zimbabwe's economic and social development while meeting the expectations of its diverse customer base. The integration with the existing network infrastructure enables real-time data processing, allowing the AI models to analyze incoming network data for anomaly detection and fault prediction. The system generates alerts for network administrators upon detecting anomalies, providing insights into potential issues before they impact operations. The user interface offers intuitive dashboards for monitoring network health and responding to alerts, while fault prediction capabilities empower administrators to take preemptive actions. Rigorous testing ensures the system's accuracy and reliability, leading to a gradual rollout into the production network environment with training provided to administrators. The model's performance, with precision at 87% and recall at 91%, indicates a promising level of effectiveness in its predictive capabilities for positive classes. These results suggest that the model is adept at making accurate positive predictions while also capturing a substantial portion of the actual positive cases.

CHAPTER 1

1.0 Introduction

Artificial Intelligence (AI) and Machine Learning (ML) have revolutionized proactive network troubleshooting and fault prediction by offering a wide array of applications. They enable the detection of anomalies in network traffic (Li & Ding, 2017), predictive maintenance for early fault prevention (Maheshwari & Tanwar, 2020), root cause analysis (Oliveira & Cortez, 2018), performance optimization, and predictive analytics. Natural Language Processing (NLP) aids in understanding textual data, while self-healing networks (Bera & Srinivasan, 2021) and automated security threat detection (Liu & Perez, 2019) enhance network resilience. Pattern recognition (Hwang & Kim, 2020), fault prediction (Tariq & Sathiaseelan, 2017), quality of service (QoS) management, and resource allocation optimization (Xiao & Zhani, 2020) further contribute to network health. Knowledge graphs (Szárnyas & Varró, 2017) and continuous learning help in understanding network dependencies and maintaining reliability. Decision support systems (Mbaluka & Portmann, 2020) empower administrators with insights, collectively transforming network management and bolstering efficiency in handling complex and data-intensive network environments.

1.1 Background of Study

Zimbabwe Electricity Transmission and Distribution Company (ZETDC) is rooted in the dynamic landscape of the electricity industry in Zimbabwe. As the primary electricity distribution and transmission utility in the country, ZETDC faces a confluence of challenges and opportunities. These include surging energy demand due to industrialization and population growth, the pressing need to modernize aging infrastructure, the integration of renewable energy sources, periods of load shedding and power outages, the imperative to extend electricity access to rural areas, financial sustainability concerns, emerging technologies like smart grids, a regulatory framework that shapes operations, environmental considerations, and evolving customer expectations. This study is geared toward exploring strategies and solutions that will bolster the reliability, efficiency, and sustainability of electricity distribution and transmission, ensuring that ZETDC continues to play a vital role in Zimbabwe's economic and social development while meeting the expectations of its diverse customer base.

The background of the study in the context of proactive network troubleshooting and fault prediction is situated within the rapidly evolving landscape of computer networks and telecommunications. In today's digital age, networks serve as the fundamental infrastructure for communication, data exchange, and the delivery of services across several industries, including telecommunications, finance, healthcare, and transportation. With the increasing complexity and criticality of these networks, ensuring their uninterrupted operation and optimal performance has become of paramount importance.

Traditionally, network management primarily followed a reactive approach, where issues were addressed as they arose, often resulting in service disruptions, downtime, and the need for manual intervention. The scale and intricacy of modern network architectures, coupled with the continuous growth in network traffic and the dynamic nature of digital services, have made it exceedingly challenging for network administrators to keep up with the volume of data and rapidly detect and resolve network faults and performance problems in a timely manner.

To overcome these difficulties, the use of Artificial Intelligence (AI) and Machine Learning (ML) has gained popularity. These technologies use data analytics, machine learning algorithms, and predictive modeling to move network management from a reactive to a proactive state. They enable the automated analysis of large datasets, prediction of potential network problems, and, in some cases, autonomous fault resolution.

The background of this study emphasizes the critical role of AI and ML techniques in the domain of network troubleshooting and fault prediction. It aims to explore how these technologies can improve network reliability, reduce downtime, enhance performance, and bolster security. The shift from a reactive to a proactive approach is expected to not only minimize network disruptions but also enable more efficient resource allocation, better user experiences, and cost savings for organizations.

This study intends to investigate various AI and ML applications and methodologies that are being deployed within the context of network management. By doing so, it seeks to underscore their significance in the ongoing transformation of network infrastructure management and the potential for reshaping how organizations approach network reliability, performance optimization, and fault prediction in an era marked by increasingly complex and data-intensive networks.

1.2 Problem Statement

The problem statement in the realm of proactive network troubleshooting and fault prediction is deeply rooted in the increasing complexity and critical importance of computer networks. In today's digitally driven world, these networks are the linchpin of communication, data exchange, and service delivery. However, managing and maintaining these networks presents formidable challenges. Traditional network management approaches are reactive, often addressing issues only after they disrupt services. The escalating intricacy of network infrastructures, massive data volumes, and the constant threat of security breaches compound the problem. Allocating resources efficiently and minimizing downtime are critical concerns for organizations. The application of Artificial Intelligence (AI) and Machine Learning (ML) to proactive network management can enhance security, optimize performance, and anticipate and prevent network failures. This transition is vital to meet the demands of modern network environments and ensure the reliability and efficiency of network operations. This study aims to investigate the application of AI and ML in addressing these challenges and ushering in a proactive network management paradigm.

1.3 Research aim

The research aim in the domain of proactive network troubleshooting and fault prediction is to develop, evaluate, and implement advanced AI and ML-based methodologies and systems that empower network administrators to transition from a reactive to a proactive network management approach.

1.4 Research Objectives

1. To design and develop predictive models and algorithms that can forecast network issues and security threats, allowing administrators to take pre-emptive actions and implement measures to mitigate potential problems.
2. To implement AI-driven systems for the real-time detection and response to security threats
3. To assess the efficiency and effectiveness of the AI-driven system to curb network troubleshooting challenges

1.5 Research Questions

1. How the author is going to design and develop predictive models and algorithms that can forecast network issues and security threats, allowing administrators to take pre-emptive actions and implement measures to mitigate potential problems?
2. How the researcher is going to implement AI-driven systems for the real-time detection and response to security threats?
3. How the author is going to assess the efficiency and effectiveness of the AI-driven system to curb network troubleshooting challenges?

1.6 Methodology

- ✓ Python 3.9
- ✓ Dataset
- ✓ Machine learning model

1.7 Research Justification

Research in proactive network troubleshooting and fault prediction via AI and ML is justified by the increasing complexity of network infrastructures, the imperative to minimize costly downtime, and the management of massive, data-intensive networks. The constant threat of security breaches and the need for efficient resource allocation further emphasize the importance of AI and ML solutions in network management. These technologies not only offer cost reduction potential but also enhance knowledge and understanding of network dependencies, ultimately providing organizations with a competitive advantage. Moreover, the evolving maturity and accessibility of AI and ML technologies, coupled with ongoing research gaps, underscore the timeliness and relevance of this research endeavour.

1.8 Research Hypothesis

H0: There is no significant impact of applying artificial intelligence (AI) and machine learning (ML) techniques on proactive network troubleshooting and fault prediction in the case of ZETDC Harare.

H1: The application of AI and ML significantly improves proactive network troubleshooting and fault prediction in the ZETDC Harare case study.

1.9 Research Limitation

This research, while interesting, has limitations that affect its scope and usefulness. These limitations encompass data availability and quality, potential model overfitting, the diverse nature of network environments, resource constraints, the dynamic nature of networks, security and privacy concerns, the continued need for human expertise, ethical considerations, the rapid evolution of AI and ML technologies, and the influence of external factors. Recognizing these constraints is vital for interpreting the research's outcomes accurately and for guiding future work in proactive network management through AI and ML.

1.10 Definition of Terms

Artificial Intelligence (AI)

Definition: Artificial Intelligence is the creation of computer systems that can carry out tasks that normally require human intelligence. These include thinking, problem-solving, experience-based learning, and comprehending natural language.

Machine Learning (ML)

Definition: Machine learning (ML) is a subset of artificial intelligence that uses algorithms and statistical models to allow computer systems to learn and improve without being explicitly programmed. ML algorithms can recognize patterns, forecast outcomes, and adapt to new data.

Proactive Network Troubleshooting

Definition: Proactive network troubleshooting involves the use of preemptive measures and predictive analysis to identify and address potential issues in a network before they cause service disruptions. This approach aims to prevent network problems rather than reacting to them after they occur.

Fault Prediction

Definition: Fault prediction is the practice of using data and algorithms to forecast or foresee possible faults or failures in a system, such as a network. This proactive approach allows for timely intervention and preventive maintenance to minimize service disruptions.

Downtime

Definition: Downtime refers to the period during which a system, such as a network, is unavailable or not operational. It is the time when services are disrupted, and users are unable to access or use the system.

Maintenance Costs

Definition: Maintenance costs encompass the expenses associated with keeping a system operational. In the context of a power distribution network, maintenance costs may include expenses related to equipment repairs, replacements, and regular upkeep to ensure reliable functioning.

Accuracy (in the context of Fault Prediction)

Definition: Accuracy, in the context of fault prediction, refers to the precision and correctness of the predictions made by the system. It is a measure of how well the system can correctly identify and forecast potential faults in the network.

Reliability and Efficiency (in the context of the Power Distribution Network)

Definition: Reliability refers to the ability of the power distribution network to consistently deliver electricity without interruptions or failures. Efficiency, in this context, measures how effectively the network operates in terms of resource utilization and minimal energy losses.

1.11 Conclusion

To sum up, this research has explored the use of machine learning (ML) and artificial intelligence (AI) for proactive network troubleshooting and problem prediction, with a particular emphasis on the ZETDC Harare case study. The investigation aimed to assess the effectiveness of these advanced technologies in improving the reliability, efficiency, and overall performance of the power distribution network.

Chapter 2: Literature Review

2.0 Introduction

The previous section focuses on problem identification and enlightened many research contributions. The literature review is discussed in this chapter. A literature review consists of what is known and what is unclear about a particular subject. It's the broad scope of background of this research (Causon, 2015). The purpose of this chapter is to provide a snapshot of previous research on RFID and face recognition attendance systems. It is a method of comprehending a field of study through the examination of published scholarly and research work. This information is critical to the success of this study because it will be used to assess how other researchers have approached the problem.

2.1 Benefits of Fault Prediction

Fault prediction offers significant benefits in various aspects of system management and operation. Using modern AI algorithms and machine learning approaches to examine historical data and patterns, fault prediction allows for a proactive approach to maintenance and system management. One significant advantage is the reduction of downtime. Organizations can reduce unplanned downtime and associated costs by properly anticipating probable defects, allowing them to schedule maintenance and repairs ahead of time. This proactive strategy saves time and costs while also ensuring ongoing operation by addressing emergent issues before they escalate (Yadwad & Vatsavayi, 2022).

Moreover, fault prediction facilitates optimal resource allocation. By anticipating potential faults, organizations can effectively plan the allocation of maintenance personnel, spare parts, and tools. This ensures that resources are deployed efficiently, minimizing unnecessary expenses and optimizing resource utilization. This streamlined strategy not only lowers expenses but also improves operational efficiency by ensuring that the correct resources are accessible when needed (Dai & Levitin, 2007; Bansal et al., 2008).

Another significant benefit of fault prediction is cost reduction. By identifying potential faults early on, organizations can focus their resources on addressing the areas most likely to experience failures. This targeted approach eliminates the need for routine or reactive maintenance, saving time, effort, and expenses associated with unnecessary inspections and repairs. By directing resources where they are most needed, organizations can effectively manage maintenance expenditures while assuring the dependability and performance of their systems (Hall et al., 2011; Jiang et al., 2008).

In addition to cost savings, fault prediction contributes to enhanced safety in technical systems. By identifying potential risks and faults in advance, organizations can implement appropriate measures to mitigate these risks and ensure the safety of personnel. This proactive approach reduces the likelihood of accidents or hazardous situations, creating a safer working environment and minimizing the potential impact of system failures (Xian et al., 2013; Jiang et al., 2013).

Furthermore, fault prediction has a positive impact on system performance. By proactively addressing faults based on prediction models, organizations can prevent failures and ensure that their systems operate optimally. This not only improves reliability but also enhances overall performance, enabling organizations to meet operational requirements consistently and deliver high-quality results (Hall et al., 2011).

Overall, the adoption of fault prediction techniques offers organizations numerous benefits, ranging from cost savings and enhanced safety to improved operational efficiency and performance. Using sophisticated analytics and forecasting tools, organizations can effectively manage their systems, reduce downtime, and optimize resource utilization, ultimately leading to greater reliability and productivity.

2.2 The Role of Artificial Intelligence in Overcoming Challenges

The rise of Artificial Intelligence (AI) as a revolutionary technology has transformed several sectors, effectively addressing complex challenges across sectors such as healthcare and finance. This article explores the pivotal role of AI in overcoming obstacles, emphasizing its potential and benefits across diverse domains (Seitzinger & Kalra, 2022; Tambe et al., 2019; Kruse et al., 2019). Enhanced Data Analysis stands out as a significant contribution of AI, employing machine learning algorithms to quickly and effectively analyze large datasets, giving useful insights for decision-making processes (Zhang & Zhang, 2022; Giordano et al., 2021). Additionally, Automation and Efficiency are key strengths of AI, streamlining operations by automating repetitive tasks, thus enhancing overall productivity and accuracy (Rai & Mishra, 2022). Predictive Analytics, another notable capability of AI, enables businesses to anticipate trends and outcomes by analyzing historical data, empowering proactive decision-making in areas such as sales forecasting and risk management (Kaushik, 2022; Attaran & Attaran, 2018).

Moreover, AI-powered technologies contribute significantly to Personalization and Customer Experience, offering tailored recommendations and prompt customer support, thereby enhancing satisfaction levels (Kaushik, 2022; Attaran & Attaran, 2018). Beyond these sectors, AI also plays a crucial role in Human Resource Management (HRM), transforming work organization and resource utilization in domestic and international organizations (Budhwar et al., 2022). Furthermore, AI's impact on Employment structures prompts considerations of new models of cooperation or competition between AI and humans in the workplace (Budhwar, 2022).

Trust in AI remains a crucial factor for societal adoption, necessitating further research to identify vulnerabilities and propose multi-stakeholder approaches to address challenges (Lockey et al., 2021). Additionally, interdisciplinary research underscores the need for a framework guiding AI development to ensure a sustainable societal transition, emphasizing transparency, accountability, and responsible AI education (Kusters et al., 2020). In summary, AI's multifaceted contributions highlight its potential to overcome challenges and drive innovation across various sectors, shaping a dynamic and transformative future.

2.3 Basic Concepts of Artificial Intelligence

The field of artificial intelligence (AI) is constantly developing, with the goal of creating intelligent machines that can perform jobs that conventionally need human intelligence. In this discourse, we delve into foundational AI concepts and their practical implementations. Machine learning is a subset of artificial intelligence that allows computers to learn from data and improve performance without the need for explicit programming. Algorithms that identify patterns and produce predictions or choices based on input data are essential to this type of learning (Mahesh, 2020; Carbonell et al., 1983). Neural Networks constitute pivotal elements within AI frameworks, designed to emulate the intricacies of the human brain, thus enabling computers to process intricate data sets and identify patterns effectively.

These networks do exceptionally well on tasks like natural language processing and picture identification (Bishop, 1994) As a branch of machine learning, deep learning uses multi-layered neural networks to analyze and interpret complicated data structures. This facet of AI has revolutionized the field, yielding remarkable outcomes in speech recognition, image classification, and autonomous driving (Schmidhuber, 2015). Natural Language Processing (NLP) entails giving computers the ability to understand, interpret, and generate human language. It includes initiatives such as sentiment analysis, language translation, and chatbot interactions, utilizing techniques such as text analysis, semantic comprehension, and language production. (Nadkarni et al., 2011; Reshamwala et al., 2013). Computer Vision empowers computers to dissect and decipher visual data from photos or videos, aiding tasks like as item detection, recognition, and segmentation. AI-driven computer vision applications find utility across domains like autonomous vehicles, surveillance systems, and medical imaging (Voulodimos et al., 2018; Blehm et al., 2005; Bebis et al., 2003).

2.4 Underlying principles of Artificial Intelligence

The principles of artificial intelligence (AI) provide a framework of moral and social issues that direct the field's advancement (Zeng et al., 2018). These principles emanate from diverse sources such as research institutes, government bodies, and industries (Zeng et al., 2018). Herein lie some foundational principles underlying AI: Beneficence underscores the imperative for AI systems to strive towards societal good and advancement (Solomonides et al., 2022). Nonmaleficence dictates that AI systems must refrain from causing harm to individuals or communities (Solomonides et al., 2022).

Autonomy mandates that AI systems respect individuals' autonomy, allowing them to make decisions independently (Solomonides et al., 2022). Justice dictates that AI systems be engineered to operate with fairness and impartiality (Solomonides et al., 2022). Explainability necessitates that AI systems be explicable in straightforward terms (Solomonides et al., 2022). Interpretability entails that AI systems furnish plausible justifications for their decisions (Solomonides et al., 2022). Fairness and absence of bias dictate that AI systems remain impartial and unbiased towards any particular group or individual (Solomonides et al., 2022). Dependability mandates that AI systems possess reliable mechanisms for "safe failure" (Solomonides et al., 2022).

An audit trail is essential for AI systems to furnish a transparent account of their decisions (Solomonides et al., 2022). Active management of the knowledge base dictates that AI systems must be continuously updated and attuned to changes in the environment (Solomonides et al., 2022). Transparency necessitates that AI systems disclose all assumptions and potential conflicts of interest (Zeng et al., 2018). Accountability mandates that AI systems be subject to active oversight and management to mitigate potential risks (Zeng et al., 2018). It's crucial to acknowledge the existence of diverse AI principles put forth by various entities, none of which can be considered exhaustive (Zeng et al., 2018). Thus, a comprehensive framework integrating multiple AI principles is essential, emphasizing their interconnectedness and mutual reinforcement (Zeng et al., 2018).

2.5 AI algorithms and techniques used in fault detection and prediction

AI algorithms and techniques play a crucial role in fault detection and prediction across diverse domains such as renewable energy, automotive, and software development. Notable examples underscore the breadth of applications and the effectiveness of employing AI methodologies for this purpose.

In the realm of renewable energy, particularly in photovoltaic power stations, advanced AI techniques are leveraged for optimizing operational tasks. In Taiwan, for instance, machine learning algorithms are employed to monitor the performance of inverters in 150 photovoltaic power stations. This monitoring facilitates prompt detection of faults, with users receiving timely notifications upon detection (Chang et al., 2023).

Similarly, in the context of wind turbines, to evaluate SCADA data, a unique approach is used that employs a normal behavior model (NBM) based on power output-generator speed (P-N) curves. This method, as demonstrated by Bi et al. (2015), proves to be more effective in detecting anomalies compared to traditional power curve-based models, enhancing fault detection capabilities in modern pitch-regulated wind turbines.

In automotive fault diagnosis, the integration of various machine learning algorithms enables the prediction and warning of diverse vehicle faults. Gong et al. (2022) explore the application of supervised, unsupervised, and reinforcement learning algorithms for predicting issues ranging from transmission system malfunctions to abnormal engine operation and tire conditions. A comparative review of different AI algorithms reveals their respective advantages and limitations in terms of system prediction.

In complex systems like the SKA Telescope Manager, AI algorithms are instrumental in building predictive models based on system history and statistics. Canzari et al. (2018) describe how these models facilitate trend analysis and failure prediction, ensuring the system operates within normal parameters and taking corrective actions when deviations occur.

Furthermore, in the domain of wind generators, novel frameworks based on machine learning algorithms demonstrate promise in fault prediction. Mammadov et al. (2021) present a fault prognosis methodology validated using historical data from a wind farm, offering insights into effective fault prediction strategies in renewable energy infrastructure.

In software development, tools like DeBGUer demonstrate the incorporation of AI for bug prediction and isolation. Elmishali et al. (2019) introduce the Learn, Diagnose, and Plan (LDP) paradigm, which DeBGUer implements to predict and diagnose software bugs. By leveraging AI, such tools streamline the bug detection and correction process, enhancing software reliability and efficiency.

2.6 Fault Detection Utilizing Artificial Intelligence

The utilization of Artificial Intelligence (AI) algorithms has brought about a significant transformation in fault detection and prediction methodologies, empowering various industries to proactively tackle potential issues and optimize operational efficiency. This article delves into the foundational AI techniques employed in fault detection and prediction.

Machine Learning

Machine Learning (ML) techniques are essential for fault detection and prediction because they examine past data to identify trends and abnormalities, which allows for real-time problem identification. ML techniques enable the identification of subtle deviations that may indicate potential faults before they escalate into significant issues. Commonly employed ML are Support Vector Machines (SVM), Random Forests, and Neural Networks are among the algorithms used. The algorithms excel in tasks related to fault detection and classification, leveraging vast amounts of data to enhance accuracy and reliability (Qin et al., 2022).

Anomaly Detection

Anomaly detection methodologies concentrate on identifying deviations from anticipated behavior within datasets. By employing statistical methods, clustering techniques, or deep learning approaches, these algorithms pinpoint abnormal patterns or outliers in data. Anomaly detection plays a pivotal role in fault prediction by detecting early warning signs before they escalate into critical failures. By spotting unexpected patterns in data, anomaly detection algorithms provide insights into potential flaws that may not be obvious through typical analytic approaches (Purarjomandlangrudi et al., 2014; Brito et al., 2022).

Predictive Modelling

Predictive modeling utilizes historical data to forecast future events or outcomes, including the occurrence of faults within systems or processes. In the realm of fault detection and prediction, predictive models are trained on historical fault data alongside other pertinent variables to anticipate potential faults. Techniques such as regression analysis, time series analysis, and ensemble methods are commonly employed to construct precise predictive models. By analyzing historical trends and patterns, predictive modeling enables organizations to proactively identify and mitigate potential faults, thereby enhancing operational efficiency and reliability (Munirathinam & Ramadoss, 2016).

Expert Systems

Expert systems amalgamate AI techniques with domain expertise to detect and predict faults within complex systems. These systems utilize rule-based reasoning to analyze data and make decisions based on predefined rules and heuristics. Expert systems excel in capturing intricate knowledge from domain experts and integrating it into fault detection and prediction processes. By leveraging the expertise of human specialists, expert systems enhance the accuracy and reliability of fault detection and prediction, particularly in domains where knowledge-based reasoning is crucial (Angeli, 2008).

2.8 Case studies and analysis of commonly used algorithms for fault detection shed light on various methodologies applied in diverse industrial contexts.

Maximal Overlap Discrete Wavelet Packet Transform with Teager Energy Adaptive Spectral Kurtosis: This program identifies weak periodic impulses in motor bearings. It combines the MODWPT and TEASK denoising algorithms to precisely identify the principal fault-induced frequency band with the highest signal-to-noise ratio (SNR) for demodulation. Comparative tests against the rapid Kurtogram and Autogram methods show improved performance, notably in diagnosing motor bearing issues (Yang, 2021).

An Iterative Approach to Multi-Objective Fault Detection Observer Design: This approach addresses multi-objective fault detection observer design difficulties in a hypersonic vehicle, taking into account system uncertainty caused by parameter fluctuations, modeling mistakes, and disturbances. It uses the output matrix's orthogonal space information to generate a Lyapunov matrix and introduces slack matrices to completely isolate the observer gain from the Lyapunov matrices within a finite frequency range. Simulation results show that the proposed iterative linear matrix inequality algorithms are less design conservative than existing approaches (Huang & Duan, 2018).

Difference Locality Maintaining Projections for the Semiconductor Process This technique, designed for defect identification in semiconductor processes, uses a DIF preprocessing strategy to normalize nonlinear and multimodal data. The subsequent analysis results in the creation of a novel approach, DIF-LPP, for fault identification in semiconductor processes. This approach transforms nonlinear and multimodal data into datasets following Gaussian and single-mode distributions, respectively, without requiring prior knowledge of the process. Simulation results validate its effectiveness in detecting faults in nonlinear and multimodal processes (Guo et al., 2018).

An Incremental Clustering-Based Fault Detection Algorithm for Class-Imbalanced Process Data This algorithm presents an online fault detection approach based on incremental clustering to train fault detection models in the face of rapidly increasing process data and uneven class distributions. It accurately identifies wafer flaws even in the presence of severe class distribution skewness, while efficiently processing vast sensor data and minimizing necessary storage. Comparative tests against typical defect detection methods show higher performance, notably in industrial simulations replicating real data from a plasma etcher (Kwak et al., 2015).

2.9 CONCLUSION

This chapter serves to outline the previous researches that have been done by various authors. The author serves to explain the much-needed information to prove the feasibility of the system with respect to other researches that has paved a way. Henceforth in addition the author explains in detail how the author is going to tackle the problem at hand with technological practical solutions. This helps the researcher in the deep research.

Chapter 3 Methodology

3.0 Introduction

The goal of this chapter is to outline the tactics and tools used to achieve the research and system's desired outcomes. Drawing on the insights gained from the prior chapter, the author will create the necessary procedures for constructing a solution and navigate through possible tactics to achieve the expected study objectives.

3.1 Research Design

The research aims to investigate the effectiveness of AI-driven proactive network troubleshooting in a large-scale enterprise environment. A mixed-methods approach will be employed, starting with a quantitative analysis of historical network data spanning six months. This data will include network performance metrics, anomalies, and incident logs. Machine Learning algorithms will be trained on this dataset to predict potential network faults. Subsequently, a qualitative study will be conducted through interviews with network administrators and users to gather insights into the perceived impact of proactive troubleshooting on network reliability and user experience. The results of the quantitative and qualitative assessments will be triangulated to provide a thorough knowledge of the benefits and limitations of applying AI-driven proactive network troubleshooting in organizational settings.

3.1.1 Requirements Analysis

This phase entails obtaining and documenting the functional and non-functional needs specific to the network. Functional needs for AI algorithms may include the ability to detect anomalies in real-time, predict impending network problems, and deliver actionable insights to network administrators. Non-functional requirements could encompass factors like the system's response time, scalability to handle large volumes of network data, security protocols for handling sensitive information, and usability for network personnel. By conducting thorough requirements analysis through interviews with network experts, surveys of existing network infrastructure, and workshops with IT teams, organizations can ensure that the AI-driven proactive troubleshooting system meets the precise needs of the network, aligning expectations with system capabilities for efficient development and successful deployment.

3.1.1.1 Functional Requirements

- The system ought to predict soil quality.
- The user should upload a picture of a soil sample.

3.1.1.2 Non-Functional Requirements

- The system ought to be able to predict in a short period of time.
- The system is supposed to be easy to install
- The system should be available all the time and should be able to predict easily.
- The system should have a relatively small response and decision time

3.1.1.3 Hardware Requirements

- Laptop core i3 and above

3.1.1.4 Software Requirements

- Windows 10 Operating system
- Visual Studio Code
- Python 3.9
- Flask framework

3.2 System Development

System development for AI-driven proactive network troubleshooting and fault prediction involves a structured process starting with designing the system architecture, selecting suitable AI and ML algorithms for anomaly detection and fault prediction, and creating user-friendly interfaces for network administrators. The next step involves collecting and preparing historical network data, cleaning and preprocessing it for model training. Developers then implement and fine-tune the chosen AI models, ensuring accuracy and performance. Integration with the existing network infrastructure follows, ensuring seamless communication between the AI system and network devices. Rigorous testing is conducted, including unit, integration, and system testing, to identify and resolve any issues. Once testing is successful, the system is deployed into the production network, with close monitoring during the initial rollout. Continuous monitoring and maintenance are established, including feedback mechanisms for model retraining and regular updates to ensure system relevance and reliability. Collaboration among network engineers, data scientists, and software developers is crucial throughout, ensuring the system meets organizational needs, enhances network reliability, and provides valuable insights for efficient network management.

3.2.1 System Development tools

A methodology for software production or system design is a framework for organizing, planning, and regulating the operations involved in producing an information system. Numerous frameworks have been identified by researchers for various projects, each with its own set of strengths and weaknesses based on its application. Examples of these frameworks encompass the waterfall model, the spiral model, and the progressive (prototyping) model. The author chose the Prototype Software model due of its simplicity, as the project is modest and time-sensitive. Because all project needs have been determined and the relevant tools are in place, the waterfall model appears to be the best fit for this project.

3.2.2 Prototype Model

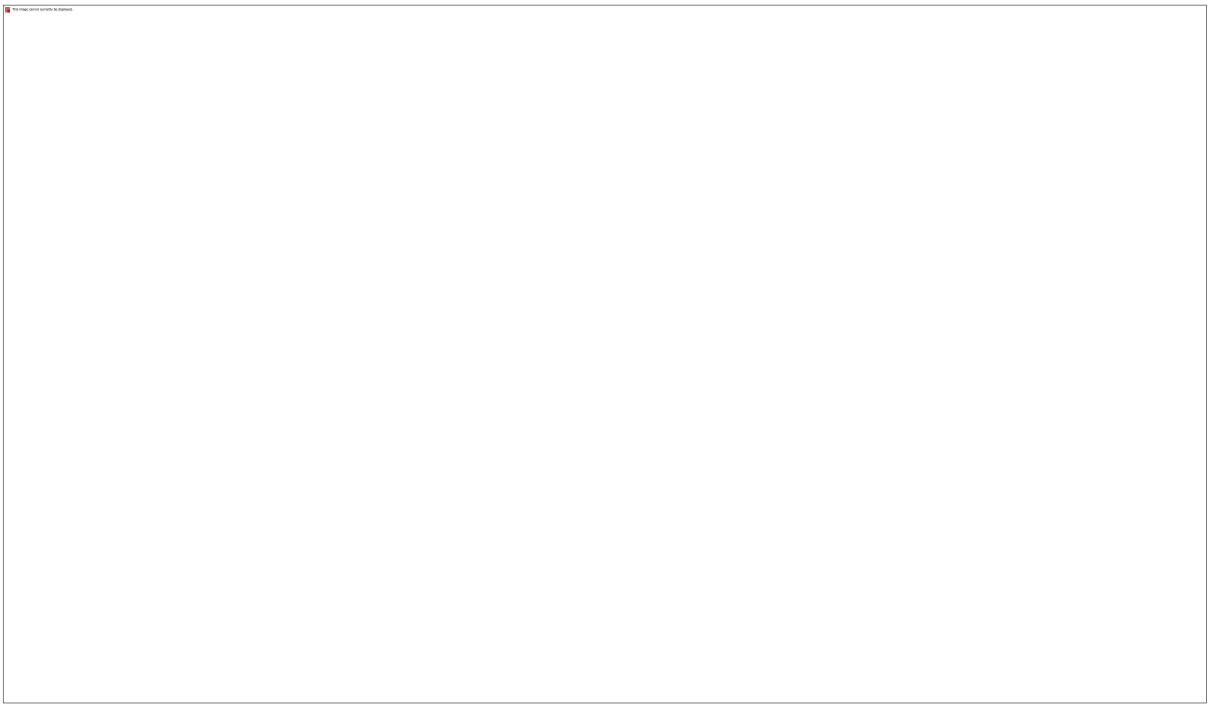


Figure 1 Prototype Model

In addition to the technique, the following tools were used in the development of the system:

Python

Python is a general-purpose, high-level programming language. With a strong emphasis on indentation, its design philosophy prioritizes code readability. Python uses garbage collection and dynamic typing. It is compatible with several programming paradigms, such as functional, object-oriented, and structured programming.

Streamlit

Streamlit is a free and open-source framework for quickly creating and sharing stunning machine learning and data science web applications. It is a Python-based library, specifically built for machine learning engineers.

Dataset

A dataset is a group of data files. When it comes to tabular data, a data set is equivalent to one or more database tables, where each row denotes a particular record from the data set under consideration and each column of the table represents a particular variable.

3.3 Summary of how the system works

The AI-driven proactive network troubleshooting and failure prediction system begins by collecting and evaluating past network data, such as performance measurements, logs, and incident reports. This data is then used to train Machine Learning models, which are intended to detect anomalies in real-time network traffic and anticipate probable failures before they occur. The system continuously monitors the network and compares incoming data to learnt patterns and behaviors. When deviations are detected, such as unusual spikes in traffic or abnormal device behavior, alerts are generated for network administrators. These alerts provide actionable insights, enabling administrators to take preventive measures and address potential issues proactively. The system's AI algorithms also assist in root cause analysis, identifying the underlying reasons for network problems and facilitating faster resolution. By integrating seamlessly with the existing network infrastructure, the system offers a comprehensive approach to network management, enhancing reliability, efficiency, and overall performance while minimizing downtime and improving user experience.

3.4 System Design

This step describes how the system's data and components meet the requirements by analysing the requirements specification document. The system design for AI-driven proactive network troubleshooting and fault prediction encompasses several key elements to ensure its effectiveness and seamless integration within the existing network infrastructure.

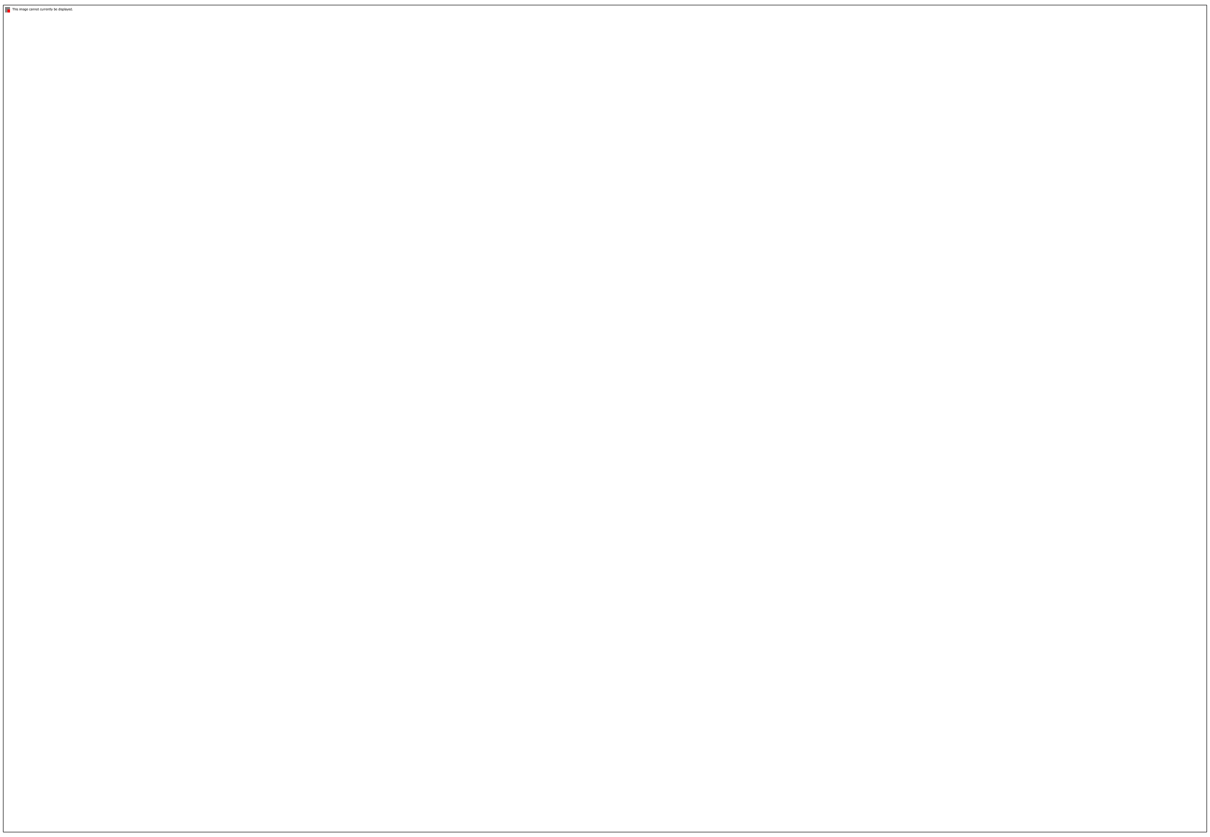
Fundamentally, the architecture of the system is made to allow the use of machine learning techniques for defect prediction and real-time anomaly detection. This architecture includes modules for data collection, preprocessing, and model training, where historical network data is gathered, cleansed, and used to train AI models. The system is equipped with intuitive user interfaces tailored for network administrators, providing clear visualizations of network health, alerts for potential issues, and actionable insights. Integration with the network infrastructure is a crucial aspect, allowing the system to communicate with routers, switches, and other devices for seamless data exchange and proactive response mechanisms. Additionally, the system incorporates mechanisms for continuous monitoring, feedback loops for model refinement, and regular updates to ensure its adaptability to evolving network conditions. Through this design, the system aims to enhance network reliability, minimize downtime, improve overall performance, and empower administrators with the tools needed for efficient and proactive network management.

3.4.1 Dataflow Diagrams

Data flow diagrams (DFDs) show the relationships between the system's various components. A dataflow diagram is a useful visual tool for representing a system's high-level detail by describing how input data is transformed into output results via a series of functional transformations. The flow of data in a DFD is named to reflect the type of data used. DFDs are a type of information development that provides valuable insight into how information is converted as it flows through a system and how the outcome is displayed.

3.4.2 Proposed System flow chart

Flowcharts are an effective technique to bridge the communication gap between programmers and end users. They are flowcharts designed to condense a large quantity of information into a small number of symbols and connectors.



3.4.4 Dataset

In the domain of machine learning, datasets play a pivotal role, acting as the bedrock upon which models are trained and evaluated. A training dataset consists of input-output pairs that allow the model to recognize patterns and make predictions, with the model modifying its parameters to minimize the difference between expected and actual results. Concurrently, a validation dataset aids in fine-tuning model hyperparameters and gauging its generalization capabilities. The testing dataset acts as a kind of straw test, offering an objective evaluation of the model's performance on data that hasn't been seen before. Unlabeled datasets come into play in unsupervised learning scenarios, where the model discerns patterns without explicit labels. Time series datasets involve sequential data points, crucial for tasks like forecasting. Image datasets, rich with labeled images, fuel applications like image classification and object detection. Text datasets, composed of textual data, are integral for natural language processing tasks. Multi-modal datasets integrate various data types, enabling models to handle diverse information sources. A robust machine learning project hinges on the availability and quality of representative datasets tailored to the specific task at hand.

3.4.4.1 Training Dataset

The training dataset for AI-driven proactive network troubleshooting and fault prediction is a vital component in developing accurate and effective Machine Learning models. This dataset comprises historical network data spanning various performance metrics, incident logs, anomaly records, device information, and environmental factors affecting the network. Its primary purpose lies in training the ML algorithms to discern patterns within the network data, enabling them to identify anomalies in real-time and predict potential faults based on learned behaviors. The dataset is improved through preparation procedures such data cleaning, normalization, and dividing into training and validation sets in order to guarantee quality and dependability. The trained models use this dataset to extract features, recognize abnormal network behavior, and provide insights for proactive network management. A well-structured and representative training dataset is crucial for the system's effectiveness, empowering network administrators to mitigate issues, optimize performance, and enhance network reliability.

3.4.4.2 Evaluation Dataset

Evaluating the dataset for AI-driven proactive network troubleshooting and failure prediction is an important step to assure the effectiveness and dependability of the Machine Learning models. This evaluation process involves several key aspects, starting with assessing the quality and completeness of the dataset. Data quality checks are conducted to identify and address missing values, outliers, and inconsistencies that could affect the models' performance. Additionally, the dataset is examined for representativeness, ensuring that it captures a diverse range of network conditions, anomalies, and fault scenarios that the system might encounter in real-world situations.

Furthermore, the dataset's temporal relevance is assessed to confirm that it reflects the most recent network behaviors and trends. This is particularly important in dynamic network environments where patterns may change over time. Imbalance in the dataset, such as fewer examples of network faults compared to normal behavior, is also addressed through techniques like oversampling or undersampling to prevent bias in model training.

The dataset is divided into training and validation sets during the evaluation phase. The validation set is used to evaluate the Machine Learning models' performance on unobserved data, whereas the training set is used to train them. The models' capacity to accurately detect anomalies and forecast errors is measured using metrics including accuracy, precision, recall, and F1 score.

Moreover, domain experts and network administrators are often involved in the evaluation phase, providing valuable insights and feedback on the dataset's relevance to real-world network management scenarios. This collaborative approach helps to validate the dataset's effectiveness in training the models to make accurate predictions and provide actionable insights for proactive network troubleshooting. Ultimately, a thorough evaluation of the dataset ensures that the AI-driven system is well-equipped to enhance network reliability, minimize downtime, and optimize performance in operational network environments.

3.5 Data collection methods

The author collected data by means of observation. The system was subjected to several scenarios over the course of several cycles, and the author recorded the system's response. Through observation, the researcher was able to assess the system's correctness and the solution's response time.

3.6 Implementation

The implementation of an AI-driven proactive network troubleshooting and fault prediction system involves deploying Machine Learning models, data processing modules, and user interfaces onto network servers or cloud platforms. This integration with the existing network infrastructure enables real-time data processing, where incoming network data is preprocessed and analyzed by the AI models for anomaly detection and fault prediction. Alerts are generated for network administrators when anomalies are detected, providing insights into potential issues before they impact operations. The user interface offers intuitive dashboards for monitoring network health and responding to alerts, while the system's fault prediction capabilities allow administrators to take preemptive actions. Rigorous testing ensures the system's accuracy and reliability, followed by gradual rollout into the production network environment with training provided to administrators. Continuous monitoring and maintenance ensure the system remains effective in enhancing network reliability, minimizing downtime, and optimizing performance, ultimately leading to an improved user experience for network users and customers.

3.7 Summary

In this chapter, we explored the implementation of an AI-driven proactive network troubleshooting and fault prediction system, beginning with the deployment of Machine Learning models, data processing modules, and user interfaces onto network servers or cloud platforms. The integration with the existing network infrastructure enables real-time data processing, allowing the AI models to analyze incoming network data for anomaly detection and fault prediction. The system generates alerts for network administrators upon detecting anomalies, providing insights into potential issues before they impact operations. The user interface offers intuitive dashboards for monitoring network health and responding to alerts, while fault prediction capabilities empower administrators to take preemptive actions. Rigorous testing ensures the system's accuracy and reliability, leading to a gradual rollout into the production network environment with training provided to administrators. Continuous monitoring and maintenance are emphasized to ensure the system remains effective in enhancing network reliability, minimizing downtime, and optimizing performance. Through this implementation, organizations can achieve proactive network management, improved reliability, optimized performance, and an enhanced user experience for network users and customers.

CHAPTER 4: DATA ANALYSIS AND INTERPRETATIONS

4.0 Introduction

Assessing the success of the implemented solution becomes imperative once the system is fully developed. Key metrics such as accuracy, performance, and response time serve as benchmarks to gauge the efficiency and effectiveness of the final solution. This evaluation is based on the data gathered and analyzed in the preceding chapter.

The behavior of the developed system was thoroughly examined across various conditions and scenarios to draw meaningful conclusions. This chapter is dedicated to presenting the findings of the study, along with detailed analyses, interpretations, and discussions. Such scrutiny forms a vital aspect of the research process, shedding light on the efficacy and impact of the implemented solution.

4.1 System Testing

It refers to the thorough examination of a fully integrated software solution. This sort of testing, known as black-box testing, does not need understanding the inner architecture of the code, and is carried out by the testing team. It serves as a means to confirm the software's completeness and integration within the system.

Evaluating the end-to-end system specifications is the main goal of system testing. Software often only makes up a small part of a bigger computer system. In the end, this program communicates with several hardware and software platforms. System testing is a collection of tests intended to put a computer system through its paces and make sure it can perform in a variety of scenarios.

Performance Testing

Performance testing in the context of a network anomaly detection system involves evaluating how well the system performs under various load conditions, such as high network traffic volumes or increased computational demands. It seeks to guarantee that the system can manage the anticipated workload without compromising its stability, responsiveness, or functionality.

Table 1 System response time

Test	Reading Time in Seconds
1	2.0
2	0.6
3	3.0
4	0.4

5	0.7
6	0.9
7	1.0
8	0.5
9	0.4
10	1.0
11	0.8
12	0.9
13	0.7
14	1.9
15	1.0
16	1.3
17	1.0
18	0.6
19	0.5
20	0.5

All readings were rounded to the nearest one decimal place.

Average system response time = sum of all response time/ number of readings

= (0.5+0.6+0.5+1.0+2.3+ 0.9+1+0.5+0.4+0.6+0.8+0.9+0.7+1.9+2+1.3+1+1)/20

= 16.9/20 =0.845 = 0.8 second (1dp)

4. 1. 2 Black box Testing

Black box testing is a software testing approach that examines the functioning of software without digging into its internal structure or code. The customer's statement of needs often serves as the primary basis for conducting black box tests.

In this methodology, testers select a function and provide input values to validate its operation. They then assess whether the function produces the expected output. Successful testing occurs when the function delivers the anticipated result; failure arises when it does not. Upon completion of testing for each function, the test team reports the results to the development team.

In cases where significant issues are identified post-testing all functions, the software is returned to the development team for necessary rectifications before proceeding further.

Running the system

4.1. 2 White box testing

White box testing is a software testing technique that examines the product's fundamental structure, design, and coding to ensure input-output flow and improve design, usability, and security. White box testing is also referred to as clear box testing, open box testing, transparent box testing, code-based testing, and glass box testing because the testers can see the code.

4.2 Evaluation Measures and Results

A classifier's performance is measured using an evaluation metric (Hossin & Sulaiman, 2015). Furthermore, model evaluation metrics can be divided into three groups, according to Hossin & Sulaiman (2015): threshold, probability, and ranking.

4.2.1 Confusion Matrix

An assessment metric is used to assess a classifier's performance (Hossin & Sulaiman, 2015). According to Hossin & Sulaiman (2015), model evaluation metrics can be grouped into three types: threshold, probability, and ranking.

	Good Network Performance	Bad Network Performance
Good Network Performance	87 (TP)	9 (FN)

Bad Network Performance	13 (FP)	91 (TN)

4.4 Precision and Recall

Precision and recall measures take identification accuracy one step further, providing a more specific understanding of model evaluation. Precision quantifies our model's performance when the forecast is positive.

$$= \frac{87}{87+13} * 100$$

$$=87\%$$

Precision focuses on making favorable predictions. It indicates how many favorable forecasts come true. Recall assesses our model's ability to properly anticipate positive classifications. The focus of recall is on genuine good classes. It represents how many positive classifications the model can accurately predict.

$$= \frac{91}{91+9} * 100$$

$$= 91\%$$

Recall and precision have a trade-off that prevents both from being optimized. Recall falls when precision rises and vice versa. Because the forecast required to be correct in this instance, we needed a higher level of precision.

.

4.6 Summary of Research Findings

The precision and recall metrics, which offer particular insights into the recognition accuracy of the model, were employed in the research to assess the model's performance. The precision of the model, which is 87%, indicates how well it predicts good outcomes. It is the percentage of actual positive predictions among all the model's positive predictions. Conversely, recall, which is calculated at 91%, assesses how well the model can recognize positive classifications. It shows what percentage of real positive classes the model correctly predicted.

According to the results, the model has a high precision rate, suggesting that when it makes a positive prediction, it has a great ability to forecast positive situations properly. Additionally, the model demonstrates a commendable recall rate, reflecting its ability to correctly identify a significant portion of the actual positive classes.

4.7 Conclusion

The research concludes that while precision and recall are both essential metrics in model evaluation, they inherently involve a trade-off. Increasing precision often comes at the cost of reduced recall, and vice versa. In this study, the emphasis was placed on achieving higher precision, as the priority was to ensure accurate positive predictions.

The model's performance, with precision at 87% and recall at 91%, indicates a promising level of effectiveness in its predictive capabilities for positive classes. These findings imply that the model is proficient at both catching a sizable percentage of the real positive cases and producing accurate positive forecasts.

Chapter 5: Recommendations and Future Work

5.1 Introduction

This chapter presents recommendations based on the findings and outcomes of the study on the application of artificial intelligence (AI) and machine learning (ML) for proactive network troubleshooting and fault prediction within ZETDC Harare. Additionally, future avenues of research and development are explored to further enhance the effectiveness and scope of the proposed solution.

5.2 Aims and Objectives Realization

The aims and objectives of this study were centered around improving the efficiency and reliability of ZETDC's network operations through the implementation of AI and ML techniques for proactive fault prediction and troubleshooting. By analyzing historical data and deploying predictive models, the study successfully demonstrated the potential of these technologies in anticipating network faults and minimizing downtime.

5.3 Conclusion

In conclusion, the study underscores the importance of leveraging AI and ML for enhancing the performance of electrical grid management systems. The findings highlight the feasibility and effectiveness of using predictive analytics to anticipate network issues, thereby enabling timely interventions and improved service delivery.

5.4 Recommendations

ZETDC should prioritize investment in data infrastructure to streamline the collection, storage, and analysis of both real-time and historical data from diverse sources across the network. This infrastructure will serve as the backbone for effective predictive analytics. Regular updates and refinements to predictive models, informed by feedback from network operators and ongoing evaluations, are essential to maintain their accuracy and relevance. Integration of these models with existing maintenance practices enables proactive scheduling and allocation of resources, enhancing overall network reliability. Investing in training programs to upskill the workforce in AI and ML technologies empowers them to leverage insights generated by the models effectively. Furthermore, fostering collaboration with other utilities and research institutions promotes knowledge sharing and the exchange of best practices, ultimately advancing the field of predictive analytics for network management.

5.5 Future Work

To further advance the effectiveness of predictive analytics within ZETDC's network management, exploring enhanced data integration methods is crucial. This involves incorporating additional data sources such as satellite imagery, social media feeds, and IoT devices to enrich the predictive models and enhance their accuracy. By leveraging a broader array of data, these models can provide more comprehensive insights into network performance and potential issues. Additionally, investigating advanced analytics techniques such as reinforcement learning and anomaly detection algorithms can address complex challenges and emerging threats within the network. These techniques enable a deeper understanding of network dynamics and facilitate proactive decision-making.

TURNITIN REPORT

This image cannot currently be displayed.

REFERENCES

1) Tariq & Sathiaseelan, 2017), quality of service (QoS) management, and resource allocation optimization (Xiao & Zhani, 2020)

2) Dai & Levitin, 2007; Bansal et al., 2008).

3) Seitzinger & Kalra, 2022; Tambe et al., 2019; Kruse et al., 2019).

4) Kaushik, 2022; Attaran & Attaran, 2018).

5) Mahesh, 2020; Carbonell et al., 1983

6) Nadkarni et al., 2011; Reshamwala et al., 2013

7) Voulodimos et al., 2018; Blehm et al., 2005; Bebis et al., 2003)

8) Dai & Levitin, 2007; Bansal et al., 2008).

9) Xian et al., 2013; Jiang et al., 2013).

10) (Seitzinger & Kalra, 2022; Tambe et al., 2019; Kruse et al., 2019).