**Name:** Abdulrahman AbdulKareem

**Title:** Comparative Analysis of XGBoost and Random Forest Algorithms for Transformer Failure Prediction using Grid Stability Data

# CHAPTER ONE

# INTRODUCTION

## BACKGROUND OF STUDY

For decades, the heart of any power grid, transformers, hum with the vital energy that fuels our world. Yet, despite their critical role, they remain vulnerable to breakdowns that cause costly downtime and disruptions. Traditional approaches, relying on fixed schedules or reactive repairs, often prove inefficient, leading to unnecessary maintenance and compromised power supply (Tianjin da xue et al., 2018). Predictive maintenance (PdM) emerges as a game-changer, transforming equipment management from reactive to proactive, leveraging the power of several prominent mode of maintenance to analyze data, it anticipates equipment failures before they strike, enabling timely interventions and optimized maintenance schedules. (Carvalho et al., 2019).

This research delves into the exciting realm of PdM for electrical equipment using machine learning-based maintenance. Instead of directly targeting transformers, it focuses on the untapped potential of electrical grid stability data. This treasure trove of information holds hidden clues about the health of individual components like transformers. By employing powerful machine learning algorithms such as Random Forest and Extreme Gradient Boosting (XGBoost), this research deciphers these clues and identifies patterns that might foreshadow transformer failures. Analyzing metrics like voltage fluctuations, frequency deviations, and load imbalances could allow us to predict potential breakdowns before they occur. Comparing the performance of these algorithms in terms of accuracy, efficiency, and interpretability will identify the most suitable tool for this task. Successful implementation could significantly reduce downtime, lower maintenance costs, and enhance grid resilience.

Ultimately, this project represents a crucial step towards smarter, more efficient, and more reliable power grids. By ensuring the uninterrupted flow of the vital energy that fuels our world, we can contribute to a brighter future for all.

## PROBLEM STATEMENT

In the realm of power grid management, the longevity and reliability of transformers, the core components sustaining our global energy infrastructure, face susceptibility to breakdowns, leading to costly disruptions. Conventional maintenance approaches, characterized by fixed schedules or reactive repairs, prove inefficient, resulting in unnecessary downtime and compromised power supply(Tianjin da xue et al., 2018). Existing research has explored various AI algorithms for transformer PdM, including support vector machines (SVM), neural networks (NN), and decision trees (DT). However, these studies primarily focus on direct sensor data, limiting their applicability in scenarios with limited or unreliable sensor coverage.

## AIM AND OBJECTIVES

The aim of the project is to demonstrate the feasibility and advantages of utilizing grid stability data for indirect transformer failure prediction via a comparative analysis of Random Forest and Extreme Gradient Boosting (XGBoost) algorithms, ultimately aiming to establish a reliable PdM strategy for electrical grids. The objectives of the project are:

1. Collect and analyze comprehensive transformer grid stability data.
2. Thoroughly evaluate the Random Forest algorithm's performance in predicting transformer failures based on the collected data.
3. Thoroughly evaluate the Extreme Gradient Boosting (XGboost) algorithm's performance in predicting transformer failures based on the collected data.
4. Determine the most effective algorithm and create a robust predictive model.

## JUSTIFICATION

Traditional maintenance methods for transformers, relying on schedules or reactive repairs, struggle with efficiency and resource allocation(Tianjin da xue et al., 2018). Unexpected breakdowns disrupt operations and cost dearly. Other predictive maintenance with AI algorithm offers a solution, but often relies on costly, sparse sensor data. This project explores the unexplored realm of indirect prediction using grid stability data, an abundant, readily available source. By comparing powerful algorithms like Random Forest and XGBoost in terms of accuracy, efficiency, and interpretability will identify the most suitable tool for this task. Successful implementation could significantly reduce downtime, lower maintenance costs, and enhance grid resilience.

## SCOPE OF STUDY

The potential of grid stability data for indirectly predicting transformer failures is explored within this research, paving the way for a proactive and cost-effective PdM strategy. Through meticulous comparison of the performance of Random Forest and XGBoost algorithms, the hidden insights within grid metrics like voltage fluctuations, frequency deviations, and load imbalances are sought to be unlocked. This endeavor aims to enable anticipation of transformer failures before they disrupt power supply, ultimately focusing on harnessing readily available data instead of expensive and sparse sensor coverage. The potential outcome is a more resilient, efficient, and optimized grid(Çinar et al., 2020; Janiesch et al., 2021).

## DEFINITION OF TERMS

**Predictive Maintenance**

Also known as Statistical-based maintenance. It is based on the continuous monitoring of the equipment or the machine. It employs prediction tools to measure when such maintenance actions are necessary, hence the maintenance can be scheduled. Furthermore, it allows failure detection at an early stage based on the historical data by utilizing those prediction tools such as machine learning methods, integrity factors, statistical inference approaches, and engineering techniques (Çinar et al., 2020).

**Machine Learning**

Machine learning refers to the ability of systems to learn from specific training data related to a particular problem, automating the creation of analytical models and addressing associated tasks (Çinar et al., 2020; Janiesch et al., 2021).

**Electrical Grid Stability Data**

This refers to the diverse set of data collected from the electrical grid, including voltage levels, current flows, frequency variations, and other operational parameters(Breviglieri et al., 2021).

**Indirect Prediction of Transformer Failures**

Rather than directly monitoring the transformers, the project aims to use the patterns and anomalies observed in the electrical grid stability data to infer potential issues with transformers. By identifying correlations and trends in the grid data that precede transformer failures.

## PROJECT LAYOUT

The organizational structure outlined below is adhered to in the project report:

**Chapter 1:** Introduction

In this chapter, an overview of the research project is presented. It encompasses an introduction to the project, a delineation of the problems under scrutiny, the study's goals and objectives, the research's importance, the study's scope, and the layout of the project.

**Chapter 2:** Literature Review

This chapter Reviews existing PdM research for transformers, focusing on indirect sensor-based approaches and analyzing previous research utilizing AI algorithms for transformer failure prediction and identify gaps in knowledge.

**Chapter 3:** Methodology

This chapter describes the data collection process for both grid stability metrics and transformer failure records. Explaining the preprocessing techniques used to ensure data quality and prepare it for analysis. Also, specifying the evaluation metrics to be used for comparing the performance of the two algorithms.

**Chapter 4:** Result and Analysis

In this chapter, a comparative analysis of Random Forest and XGBoost is performed, highlighting their strengths and weaknesses in predicting transformer failures based on grid stability data.

**Chapter 5:** Conclusion and Recommendation

This chapter Summarizes the key findings of the project, emphasizing the feasibility and advantages of using grid stability data for indirect transformer failure prediction. Also to draw conclusions about the effectiveness of Random Forest and XGBoost in this specific context, making recommendations for future PdM applications.

# CHAPTER TWO

# LITERATURE REVIEW

## INTRODUCTION

This chapter aims to review both past and present literature pertinent to the topic of “Comparative Analysis of XGBoost and Random Forest Algorithms for Transformer Failure Prediction using Grid Stability Data”. The research process commenced with a comprehensive review of journals, magazines, books, and internet sites in the field of machine learning algorithms, specifically focusing on XGBoost and Random Forest, and their application in predicting transformer failures.

## THEORETICAL BACKGROUND

### MAINTENANCE

Like any complex machinery, transformers rely on diligent maintenance to maintain peak performance and minimize disruptive failures. According to EN 13306, maintenance encompasses a holistic approach of "technical, administrative, and managerial actions" throughout the transformer's lifespan. The maintenance of transformers in a power grid involves a set of essential activities aimed at ensuring the uninterrupted and efficient functioning of the equipment(Rojek et al., 2023). These maintenance tasks include repairing, replacing components, routine checking, adjusting parameters, testing, measuring, and fault-finding. The execution of maintenance involves implementing a defined strategy through action plans. These plans ensure adherence to guidelines, maintaining direction and facilitating subsequent activities necessary for the system's maintenance. According to standard EN 13306, these maintenance strategies can be classified in a number of ways (corrective maintenance (CM) and preventive maintenance (PM))(Coandǎ et al., 2020). The figure below shows the schematic representation of the maintenance types.

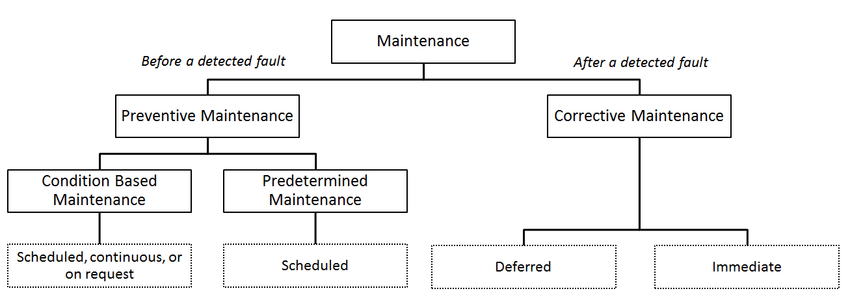


Figure 1: Types of maintenance according to EN 13306 standard

The strategies of Corrective Maintenance (CM) and Preventive Maintenance (PM) have been employed since the early 1990s. Corrective maintenance aims to restore a system after a failure, often resulting in unpredictable consequences and higher costs. Preventive maintenance involves planned interventions to keep equipment in good condition, intending to address issues before they lead to failure. However, the challenge lies in flawlessly scheduling maintenance well in advance. Increasing demands for system reliability have diminished the effectiveness of preventive maintenance, with the downside of not always considering the actual system state, leading to potential unnecessary procedures and additional costs(Coandǎ et al., 2020; Rojek et al., 2023).