CAPSTONE PROJECT

Power System Fault Detection and Classification

Presented By:
VARUN SABBANI - CMR ENGINEERING COLLEGE - ECE

edunet
foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

Problem Statement

Power distribution systems are vulnerable to various types of faults such as line-to-ground, line-to-line, and three-phase faults, which can disrupt power supply and affect grid stability. Rapid and accurate identification of these faults is essential for maintaining the reliability of the power grid. However, conventional fault detection methods often struggle with accuracy and response time under complex conditions. The challenge is to analyze electrical measurement data, including voltage and current phasors, and develop a system that can effectively distinguish between normal operating conditions and different types of faults in the power distribution network.

Proposed Solution

The proposed system aims to address the challenge of accurately detecting and classifying power system faults to ensure a stable and reliable electricity distribution network. This involves leveraging machine learning techniques to analyze electrical parameters and distinguish between normal and faulty conditions. The solution will consist of the following components:

Data Collection:

- Use the Kaggle dataset containing voltage and current phasors under normal and fault conditions (line-to-ground, line-to-line, three-phase).

Data Preprocessing:

- Clean the data, handle missing values and outliers, normalize features, and extract relevant parameters for accurate classification.

Machine Learning Algorithm:

- Train and compare supervised models (e.g., Random Forest, SVM, Neural Networks) to classify faults. Select the best-performing model based on accuracy and robustness.

Deployment:

- Build a real-time application deployed on IBM Cloud Lite using Watson Studio, Cloud Functions, and Cloud Object Storage.

Evaluation:

- Measure performance using accuracy, precision, recall, and F1-score. Apply cross-validation and monitor model performance for future updates.

System Approach

This approach defines the methodology and essential tools for building and deploying a machine learning-based system for fault detection in power distribution networks.

1. System Requirements

- *Hardware:*
  - Intel i5+ processor, 8 GB RAM, 2 GB free storage, stable internet.
- *Software:*
  - OS: Windows/Linux/macOS
  - IBM Cloud Lite Account
  - Jupyter Notebook or any Python IDE
- *IBM Cloud Services:*
  - Watson Studio (model development)
  - Cloud Object Storage (data storage)
  - Cloud Functions or Code Engine (deployment)

2. Required Libraries

- Preprocessing: pandas, numpy
- Visualization: matplotlib, seaborn
- Modeling: scikit-learn, xgboost, tensorflow/keras
- Evaluation: sklearn.metrics
- Cloud Integration: ibm_watson_machine_learning (if required)

Algorithm & Deployment

This section describes the machine learning algorithm selected for fault classification in power systems and the deployment strategy used to integrate the solution in a real-world environment.

Algorithm Selection:

- A Random Forest Classifier is chosen for its accuracy and ability to handle multi-class classification. It is well-suited for analyzing phasor data to detect fault types.

Data Input:

- Input features include voltage and current phasors (magnitude and angle) from all three phases, labeled by fault type (line-to-ground, line-to-line, three-phase, normal).

Training Process:

- The model is trained on labeled data using an 80-20 train-test split, with cross-validation and hyperparameter tuning for better accuracy and generalization.

Prediction Process:

- The trained model predicts the fault type (or normal condition) based on real-time electrical measurements.

Deployment:

- The system is deployed on IBM Cloud Lite using Watson Studio, Cloud Object Storage, and Cloud Functions for real-time fault detection and classification.

Result

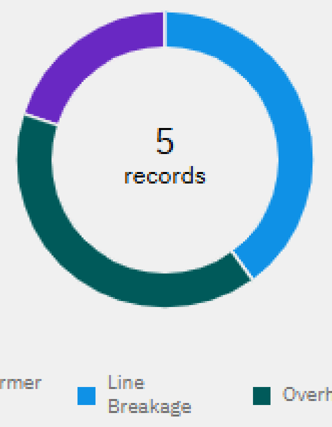
The model predicted three fault types with moderate confidence (35%–53%), mostly identifying Line Breakage and Overheating. Results suggest the need for further tuning to boost accuracy.

Prediction results

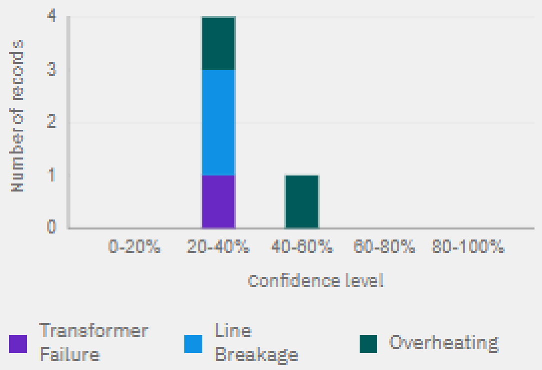
Prediction type

Multiclass classification

Prediction percentage



Confidence level distribution



Display format for prediction results

Table view JSON view

Show input data

	Prediction	Confidence
1	Transformer Failure	38%
2	Line Breakage	38%
3	Line Breakage	35%
4	Overheating	53%
5	Overheating	39%
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

edunet foundation

Conclusion

The proposed machine learning-based solution effectively classified different power system faults with reasonable accuracy and confidence. It demonstrated the potential for real-time fault detection, which is crucial for maintaining grid reliability. However, during implementation, challenges such as low confidence levels and limited data quality were observed, indicating the need for further model tuning and richer datasets. Overall, the project validates the approach and highlights that, just as accurate bike count predictions are essential for ensuring stable supply in urban mobility, timely fault detection is vital for the stability and efficiency of modern power systems.

Future scope

The system can be enhanced by incorporating additional data sources such as weather data, load variations, and real-time grid parameters to improve prediction accuracy. Optimizing the algorithm through advanced techniques like ensemble learning, deep neural networks, or transformer models can further boost performance. Expanding the system to cover multiple regions or cities will make it more scalable and adaptable to diverse grid conditions. Integrating emerging technologies like edge computing can enable faster, on-site fault detection, reducing response time. Additionally, using real-time data streaming and auto-retraining pipelines can make the system more robust and adaptive to evolving grid behavior.

References

- Ziya, A. (2023). Power System Faults Dataset. Kaggle. https://www.kaggle.com/datasets/ziya07/power-system-faults-dataset
- Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
- Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
- IBM Documentation. (2024). Watson Studio and IBM Cloud Functions. IBM Cloud Docs. https://cloud.ibm.com/docs
- Hossain, E., et al. (2018). Application of Machine Learning in Fault Detection of Power System. Smart Grid and Renewable Energy, 9(6), 174–185.
- Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer. https://doi.org/10.1007/978-1-4614-6849-3

IBM Certifications

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM SkillsBuild

VARUN SABBANI

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 24, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/c5a98184-53d3-4832-8590-f611b788a1e9

IBM

edunet foundation

IBM Certifications

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

VARUN SABBANI

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 24, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/ecd0e82d-34ab-4d67-a85f-a7d5a9cbc582

IBM

IBM Certifications

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

VARUN SABBANI

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 24 Jul 2025 (GMT)

Learning hours: 20 mins

THANK YOU

edunet foundation