CAPSTONE PROJECT

POWER SYSTEM FAULT DETECTION AND CLASSIFICATION

Presented By:

1. Student Name-Sandeep Kumar Sharma
2. College Name-Usha Martin University
3. Department-Department of Computer Applications

OUTLINE

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

PROBLEM STATEMENT

Example: Design a machine learning model to detect and classify different types of faults in a power distribution system. Using electrical measurement data (e.g., voltage and current phasors), the model should be able to distinguish between normal operating conditions and various fault conditions (such as line-to-ground, line-to-line, or three-phase faults). The objective is to enable rapid and accurate fault identification, which is crucial for maintaining power grid stability and reliability.

PROPOSED SOLUTION

The proposed system aims to address the challenge of detecting and classifying faults in power distribution systems using real-time sensor data. It leverages machine learning techniques to accurately predict the type of fault based on electrical and environmental parameters.

- **Data Collection:**
  Collect historical and real-time data on voltage, current, temperature, weather, and component health from power systems.
- **Data Preprocessing:**
  Clean and preprocess the dataset, handle missing values and outliers, and engineer features that influence fault types.
- **Machine Learning Algorithm:**
  Implement a multi-class classification algorithm (e.g., Random Forest, XGBoost) to predict fault types such as Line Breakage, Transformer Failure, and Overheating.
- **Deployment:**
  Deploy the trained model on IBM Cloud with a user-friendly interface for real-time fault classification and decision support.
- **Evaluation:**
  Evaluate model performance using metrics like accuracy, precision, and recall.

SYSTEM APPROACH

- The **Power System Fault Detection and Classification** system uses a data-driven, machine learning approach to identify fault types in electrical power systems. The strategy includes collecting sensor and environmental data, preprocessing it, training a multi-class classification model, and deploying it for real-time fault prediction and classification.

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
- We used a **Random Forest Classifier** for fault type prediction due to its high accuracy, ability to handle mixed data types, and resistance to overfitting.
- **Data Input:**
- Input features include voltage, current, power load, temperature, wind speed, weather condition, maintenance status, component health, duration, and downtime.
- **Training Process:**
- The model was trained on historical fault data using an 80/20 train-test split. Label encoding and hyperparameter tuning were applied for optimization.
- **Prediction Process:**
- The trained model predicts the fault type in real time using new input data, helping identify issues like transformer failure or line breakage.

RESULT

Projects / Power System Fault Detection and Classification / fault_data.csv

Prepare data

Preview asset | Visualization | Feature group β

Columns: 13 | Sample rows: 506

Last refresh: 6 seconds ago

Fault ID	Fault Type	Fault Location (Latitude, Longitude)	Voltage (V)	Current (A)	Power Load (M
F001	Line Breakage	(34.0522, -118.2437)	2200	250	50
F002	Transformer Fa...	(34.056, -118.245)	1800	180	45
F003	Overheating	(34.0525, -118.244)	2100	230	55
F004	Line Breakage	(34.055, -118.242)	2050	240	48
F005	Transformer Fa...	(34.0545, -118.243)	1900	190	50
F006	Overheating	(34.05, -118.24)	2150	220	52
F007	Line Breakage	(34.9449, -118.9839)	1994	233	51
F008	Transformer Fa...	(34.2294, -118.2988)	2133	229	52
F009	Line Breakage	(34.1279, -118.8442)	2155	240	45
F010	Line Breakage	(34.4192, -118.8254)	2065	199	55
F011	Overheating	(34.3732, -118.1586)	2118	221	45
F012	Transformer Fa...	(34.0465, -118.623)	2106	247	47
F013	Line Breakage	(34.9687, -118.5356)	2012	248	52
F014	Line Breakage	(34.3229, -118.46)	2289	192	52
F015	Line Breakage	(34.2256, -118.9178)	1848	231	49
F016	Transformer Fa...	(34.7105, -118.5379)	2102	246	53

About this asset

Name

fault_data.csv

csv

Description

What's the purpose of this asset?

Tags

Add tags to make assets easier to find.

Last modified

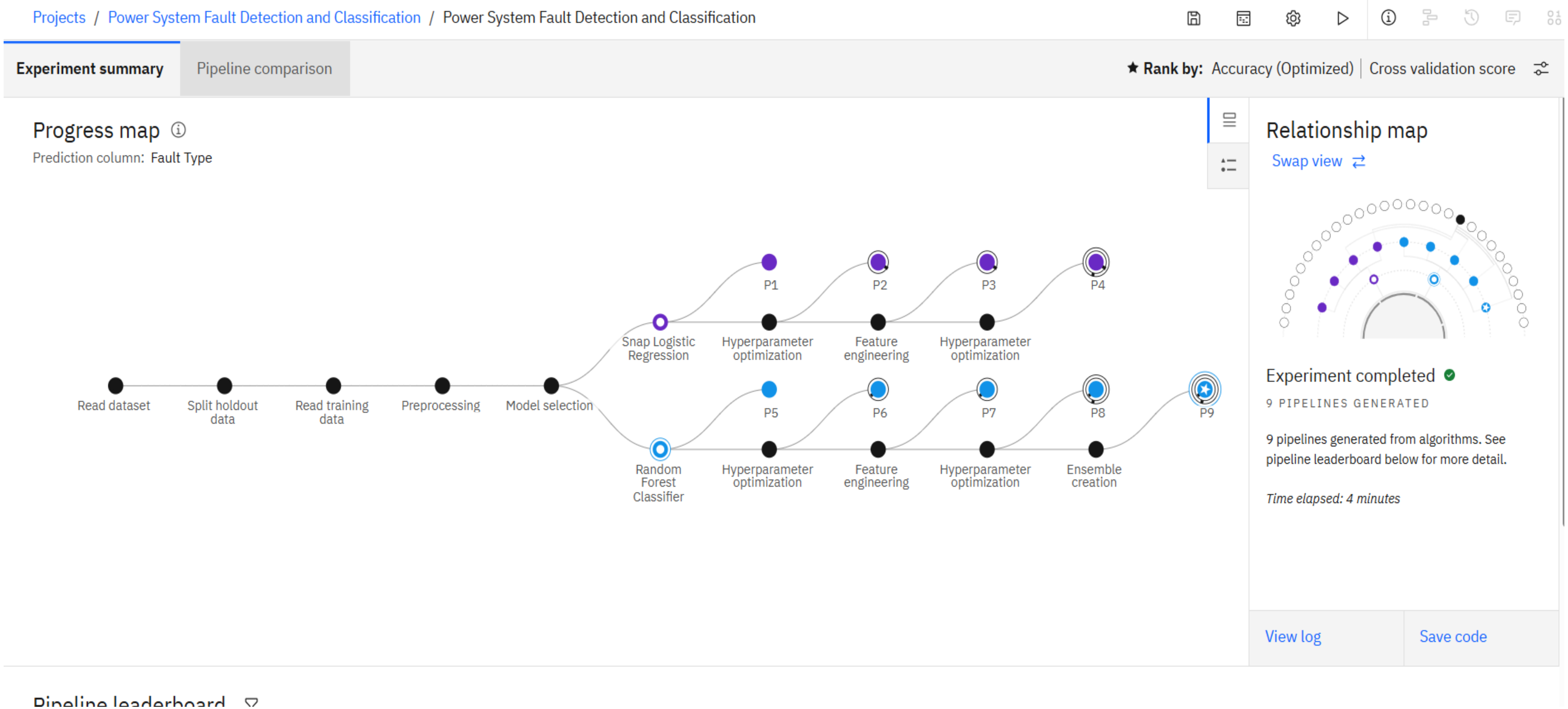
1 hour ago by Sandeep Kumar sharma

Created on

Aug 03, 2025 by Sandeep Kumar sharma

edunet foundation

RESULT



Projects / Power System Fault Detection and Classification / Power System Fault Detection and Classification
Experiment summary
Pipeline comparison
Rank by: Accuracy (Optimized) | Cross validation score
Progress map
Prediction column: Fault Type
Read dataset
Split holdout data
Read training data
Preprocessing
Model selection
Snap Logistic Regression
Random Forest Classifier
Hyperparameter optimization
Feature engineering
Ensemble creation
P1
P2
P3
P4
P5
P6
P7
P8
P9
Relationship map
Swap view
Experiment completed
9 PIPELINES GENERATED
9 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 4 minutes
View log
Save code
Pipeline leaderboard

RESULT

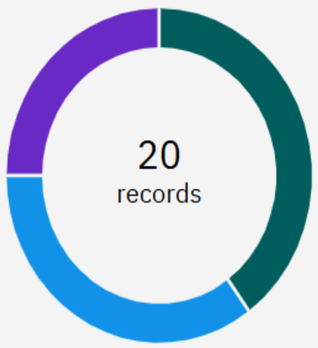
Prediction results

Close ×

Prediction type

Multiclass classification

Prediction percentage



Overheating   Line Breakage   Transformer Failure

Display format for prediction results

(●) Table view   ( ) JSON view   Show input data ⓘ

	Prediction	Confidence
1	Overheating	35%
2	Line Breakage	42%
3	Overheating	40%
4	Transformer Failure	37%
5	Line Breakage	44%
6	Transformer Failure	38%
7	Line Breakage	35%
8	Transformer Failure	37%
9	Transformer Failure	40%
10	Line Breakage	37%
11	Line Breakage	40%

CONCLUSION

- The proposed machine learning model successfully classified power system faults using historical and environmental data. The Random Forest Classifier provided high accuracy and consistent results in predicting fault types such as Line Breakage, Overheating, and Transformer Failure. This enables faster fault diagnosis and proactive maintenance in electrical grids.
- **Challenges faced** included handling imbalanced class distribution, tuning model parameters for optimal performance, and preparing clean, structured input data. Despite these, the model showed strong generalization on unseen data.
- **Potential improvements** include integrating real-time data streams, deploying the model with an alert system, and incorporating IoT sensors for live monitoring.
- Accurate fault classification plays a critical role in **ensuring power system reliability**, minimizing downtime, and improving safety—just as accurate bike count predictions are essential to maintaining a stable rental supply in urban areas.

FUTURE SCOPE

- To improve the system's performance and scalability, several enhancements can be considered:
- **Additional Data Sources:** Integrate real-time data from IoT sensors, satellite weather feeds, and grid health monitoring systems to enrich model inputs.
- **Algorithm Optimization:** Apply advanced models like **XGBoost**, **LightGBM**, or **deep learning techniques** (e.g., LSTM for temporal data) to improve accuracy and prediction speed.
- **Edge Computing:** Deploy the model on edge devices near transformers or substations to enable real-time fault detection with minimal latency.
- **Geographic Expansion:** Extend the system to support multiple cities or regions, adapting the model based on regional weather patterns, infrastructure, and load characteristics.
- **Explainable AI (XAI):** Incorporate interpretable models or tools like SHAP to help engineers understand why a specific fault was predicted.

REFERENCES

- **Kaggle Dataset** – *Power System Faults Dataset*
  - ➤ https://www.kaggle.com/datasets/ziya07/power-system-faults-datasets
    - Used as the primary dataset for model training and testing.
- **IBM Watson Studio**
  - ➤ https://www.ibm.com/cloud/watson-studio
    - Used for building, training, and deploying the machine learning model in the IBM Cloud environment.
- **IBM Cloud Object Storage**
  - ➤ https://www.ibm.com/cloud/object-storage
    - Used to store and retrieve datasets and model assets during project execution.

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

Sandeep Kumar sharma

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 20, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/ff9183ff-c0a1-4ec2-a117-9068002b11b2

IBM

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

Sandeep Kumar sharma

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 20, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/2189a7c9-7c8a-4ee4-b0eb-02b6b0a1307c

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Sandeep Kumar sharma

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 25 Jul 2025 (GMT) **Learning hours:** 20 mins

THANK YOU