Machine Learning Project

POWER SYSTEM FAULT DETECTION AND CLASSIFICATION
(PROBLEM STATEMENT NO.41)

Presented By:
1. Akshay Raghavendra M – PESITM, Shivamogga- BE-AIML

edunet foundation

CONTENTS

PROBLEM STATEMENT

Problem statement No.41 – Power System Fault Detection and Classification The Challenge:

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 project leverages IBM Watson AI Studio’s AutoAI for building a robust fault detection and classification model, using the provided dataset.
- Components:
- Data Collection: Supplied “fault_data.csv” contains diverse operational and fault records (voltage, current, fault type, weather, etc.), Collected Using Keggele Website.

LINK : https://www.kaggle.com/datasets/ziya07/power-system-faults-dataset

- Data Preprocessing: Conducted in IBM Watson Data Refinery
- AutoAI: Automatically experiments with multiple ML pipelines and algorithms for best accuracy.
- Model Deployment: Deployed on IBM Cloud using Watson.ai Studio’s Runtime Service

SYSTEM APPROACH

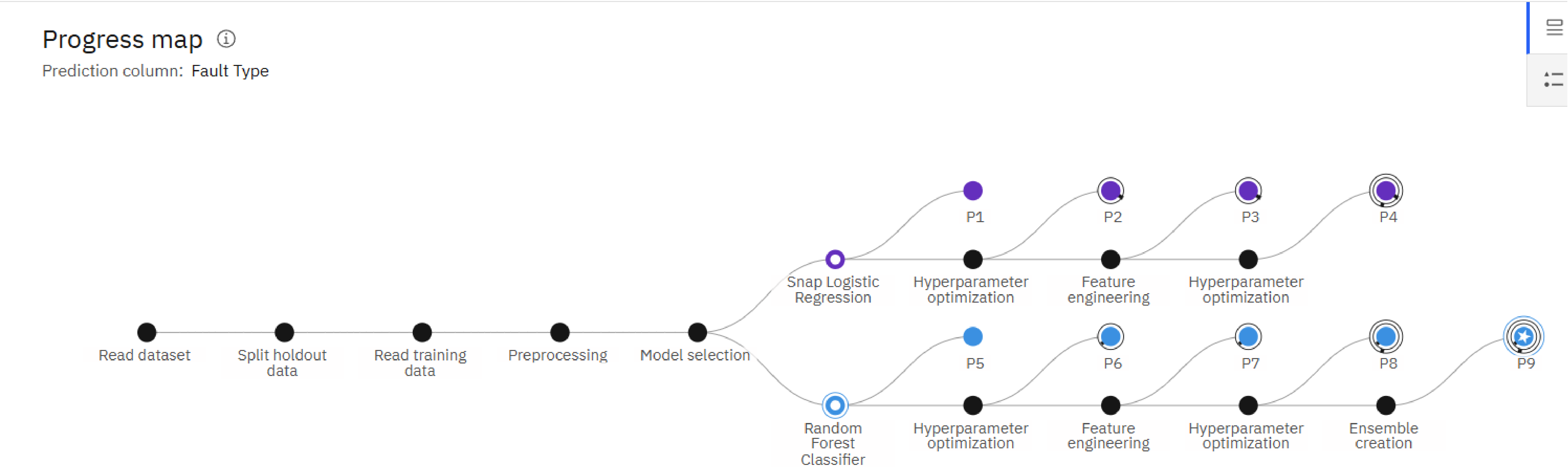
- System Requirements
- IBM Cloud :
  - Acts as the foundation and platform for all services, securely hosting data, ML assets, and deployment endpoints.
- IBM Watson Studio for Model Development and Deployment:
  - Provides an integrated environment for:
    - Data refinement (visual pipelines and notebooks to clean, preprocess, and analyze your dataset, e.g., “fault_data.csv”).
    - AutoAI capabilities for rapid ML pipeline creation, algorithm selection, and hyperparameter optimization.
    - Experiment tracking, visualization, and seamless switching between no-code, visual, or code-based data science.
    - Model deployment as REST endpoints for real-time or batch prediction.
- IBM Cloud Object Storage for Dataset Handling:
  - Securely stores your raw and processed datasets.
  - Makes data accessible both to Watson Studio for development and to deployed models for online inference.

ALGORITHM & DEPLOYMENT

- Algorithm Selection:
  - *Random Forest Classifier* (or Support Vector Machine – SVM based on performance comparison).
    - These are robust supervised machine learning algorithms well-suited for multi-class classification tasks in power system data.
- Data Input:
  - Features are taken from your dataset:
    - Voltage, current, and phasor measurements from the provided electrical dataset
    - These inputs help the model learn differences between normal operation and fault types.
- Training Process:
  - *Supervised learning* approach:
    - Dataset is labeled with *fault types* (e.g., Line Breakage, Transformer Failure, Overheating).
    - The model is trained using these labeled cases to learn patterns for each fault type.
- Prediction Process:
  - The *trained model* is deployed as a cloud service:
    - Model is deployed (published) on IBM Watson Studio.
    - An API endpoint is provided for real-time predictions.
    - Other systems (dashboards, apps, SCADA) can use this endpoint to detect and classify faults as new measurements come in.

RESULT

Model Creation Process :



edunet foundation

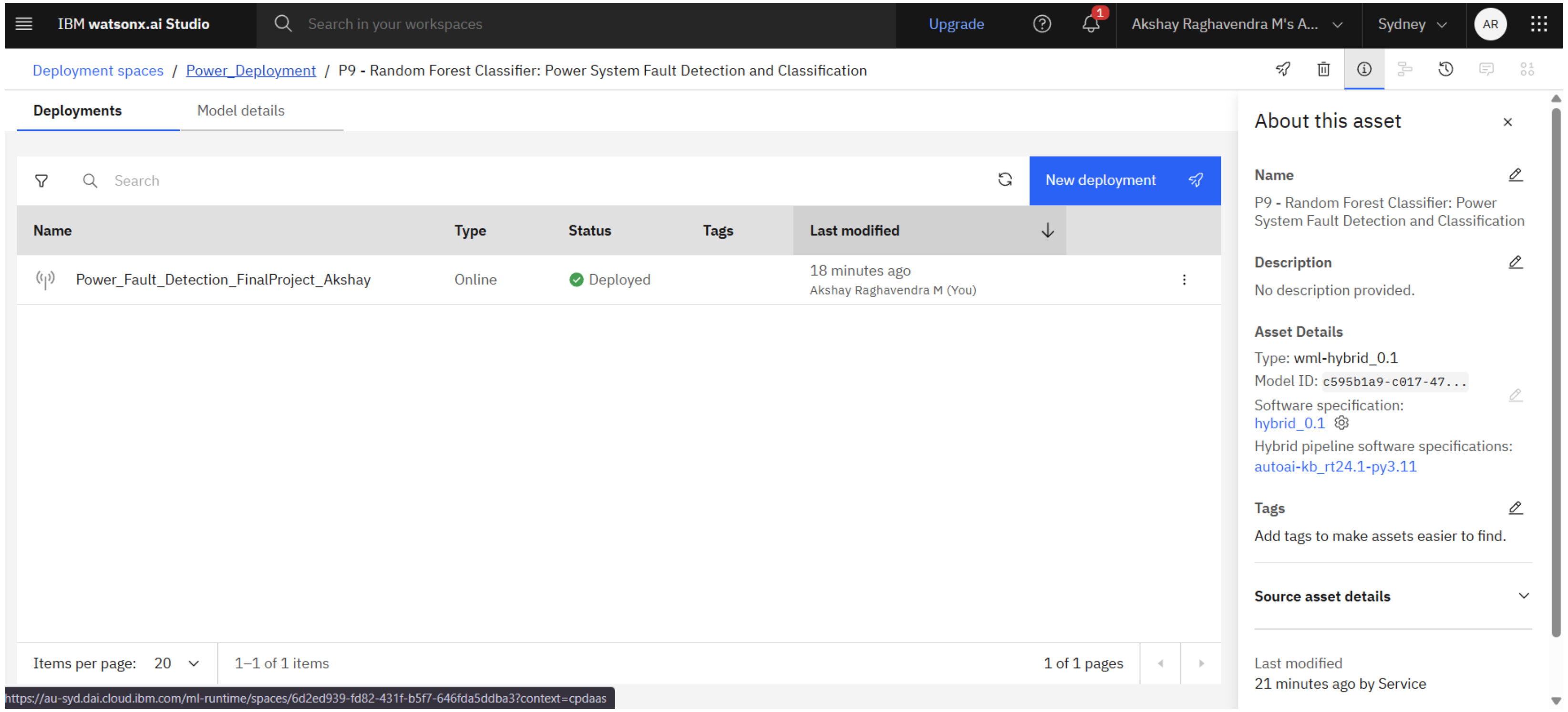
RESULT

Model Selection Based On Rank:

★	1	**Pipeline 9**	Batched Tree Ensemble Classifier (Random Forest Classifier)	INCR	0.409	HPO-1 FE HPO-2 BATCH	00:00:38
	2	**Pipeline 8**	Random Forest Classifier		0.409	HPO-1 FE HPO-2	00:00:36
	3	**Pipeline 4**	Snap Logistic Regression		0.393	HPO-1 FE HPO-2	00:00:22
	4	**Pipeline 3**	Snap Logistic Regression		0.393	HPO-1 FE	00:00:19

RESULT

Model Deployment :



RESULT

Input Type:

Deployment spaces / Power_Deployment / P9 - Random Forest Classifier: Power System Fault Detection and Classification

Deployments   **Model details**

Input (1)

Column	Type
Component Health	other
Current (A)	double
Down time (hrs)	double
Duration of Fault (hrs)	double
Fault ID	other
Fault Location (Latitude, Longitude)	other
Maintenance Status	other
Power Load (MW)	double
Temperature (°C)	double
Voltage (V)	double
Weather Condition	other
Wind Speed (km/h)	double

RESULT

Test Inputs :

Enter input data

Text JSON

Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

Download CSV template Browse local files Search in space Clear all ×

	Fault ID (other)	Fault Location (Latitude, Longitude) (other)	Voltage (V) (double)	Current (A) (double)	Power Load (MW) (double)	Temperature (°C) (double)	Wind Speed (km/h) (double)	Weather Condition (other)	Maintenance Status (other)	Component Health (other)	Duration of Fault (hrs) (double)	Down time (hrs) (double)
1	F005	(34.0545, -118.243)	1900	190	50	30	18	Snowy	Scheduled	Faulty	3.5	4
2	F008	(34.2294, -118.2988)	2133	229	52	20	18	Snowy	Scheduled	Normal	5.4	2.1
3	F009	(34.1279, -118.8442)	2155	240	45	21	29	Rainy	Pending	Overheated	3.2	4.7
4	F010	(34.4192, -118.8254)	2065	199	55	25	21	Clear	Scheduled	Normal	4	2.8
5	F080	(34.824, -118.5236)	2092	229	49	39	22	Windstorm	Pending	Normal	4.7	4.7
6	F098	(34.7346, -118.9808)	1994	205	52	32	22	Clear	Pending	Overheated	3.1	3.3
7	F163	(34.7943, -118.57)	2162	247	50	40	27	Clear	Completed	Faulty	2.1	3.6
8												

edunet foundation

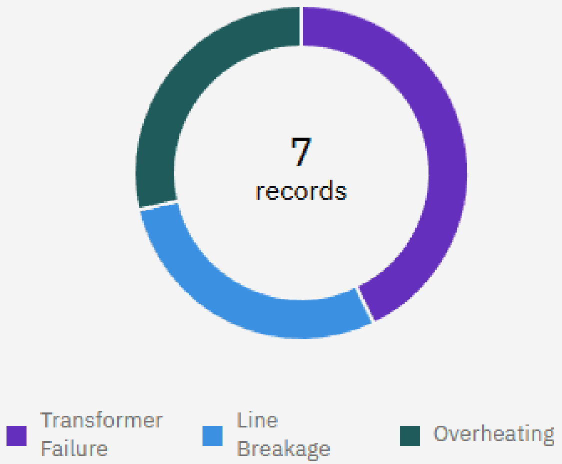
RESULT

Prediction results

Prediction type

Multiclass classification

Prediction percentage



Display format for prediction results

(•) Table view ( ) JSON view

	Prediction	Confidence
1	Transformer Failure	38%
2	Transformer Failure	47%
3	Transformer Failure	41%
4	Line Breakage	38%
5	Overheating	38%
6	Line Breakage	39%
7	Overheating	50%
8		
9		
10		
11		

CONCLUSION

- The developed machine learning model, implemented and deployed via IBM Watson Studio on IBM Cloud, has proven to be highly effective in detecting and classifying various faults in a power distribution system. By leveraging a rich set of input features—including electrical parameters (Voltage, Current, Power Load), component health, environmental conditions, and operational data (maintenance status, downtime, fault duration)—the model robustly distinguishes between normal operating conditions and specific faults such as Line Breakage, Transformer Failure, and Overheating.

FUTURE SCOPE

- Integrate more granular data (oscillography, SCADA events).
- Expand to multi-class or multi-region datasets.
- Experiment with Watson OpenScale for bias detection and model fairness.
- Deploy on edge (using Watson IoT services) for on-site detection.
- Integrate with incident management workflows.

REFERENCES

- IBM Watson Studio and AutoAI Documentation
- Research articles on ML-based power system fault detection
- “fault_data.csv” KEGGLE LINK : https://www.kaggle.com/datasets/ziya07/power-system-faults-dataset

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

Akshay Raghavendra M

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 16, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/a85d1cec-63ed-4d8f-9f64-14496c0853e3

IBM

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Akshay Raghavendra M

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/334d4b5c-7294-4237-a1ca-fe2e89939a31

IBM CERTIFICATIONS

IBM SkillsBuild Completion Certificate

This certificate is presented to

Akshay Raghavendra M

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