CAPSTONE PROJECT

POWER SYSTEM FAULT DETECTION AND CLASSIFICATION

Presented By:
Sahaja Pallapothula-IIIT Kottayam-Computer science

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

- 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

- To build a machine learning model that accurately identifies and classifies different types of faults in a power distribution network using electrical measurements like current and voltage phasors.
- Key components:
- **Data Collection:** Acquire datasets with labeled samples of fault and normal events, containing voltage and current readings from various locations in the network.
- **Data Preprocessing:** Normalize, filter noise, and transform phasor data into useful features like magnitude, angle, symmetry components, etc.
- **Machine Learning Algorithm:** Using IBM Cloud, models like Random Forest or LSTM are trained on voltage and current phasor features to classify faults (LG, LL, LLG, LLL, Normal).
- **Evaluation:** The model is trained on 70–80% of the data and evaluated using standard metrics like accuracy and precision.

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the power system fault detection and classification. Here’s a suggested structure for this section:

- System requirements
- IBM Cloud
- IBM Watson studio for model developments and deployment
- IBM cloud object storage for dataset handling

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**

  Random Forest Classifier (or SVM for linearly separable data)

- **Data Input:**

  Voltage, current and phasor measurements from the dataset

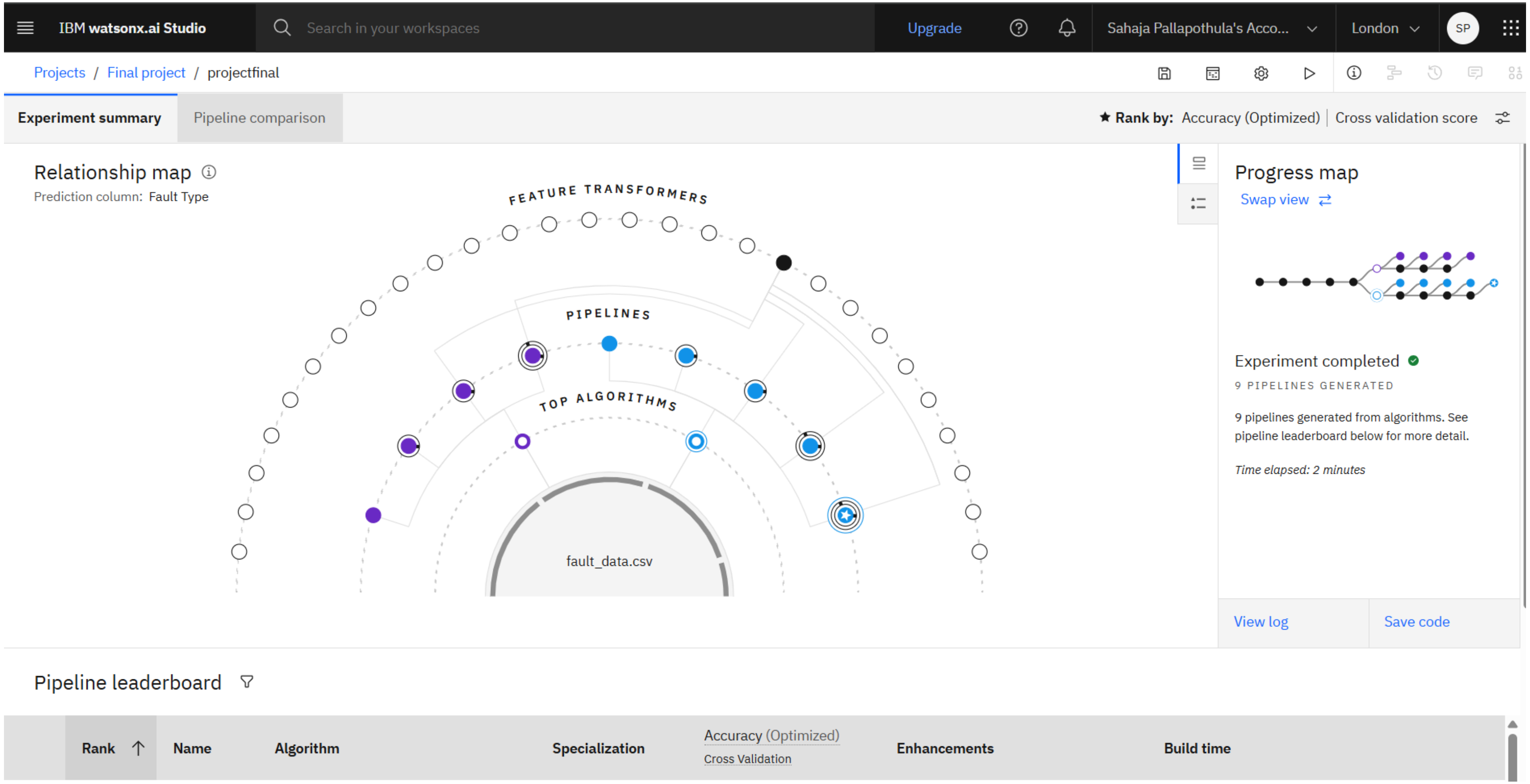
Training Process:

Supervised learning using predefined fault categories.

- **Prediction Process:**

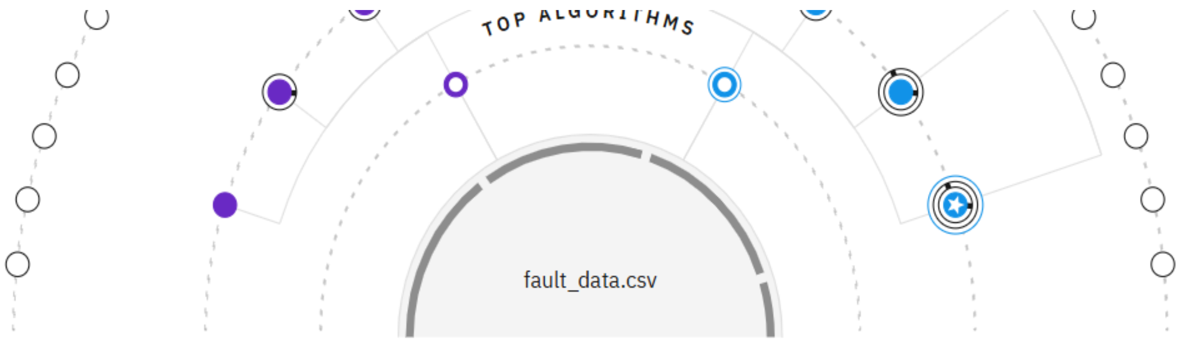
  Deployed to IBM Cloud using Watson Studio. Real-time input from sensors can trigger prediction

RESULT



IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Sahaja Pallapothula's Acco...
London
SP
Projects / Final project / projectfinal
Experiment summary
Pipeline comparison
Rank by: Accuracy (Optimized) | Cross validation score
Relationship map
Prediction column: Fault Type
FEATURE TRANSFORMERS
PIPELINES
TOP ALGORITHMS
fault_data.csv
Progress map
Swap view
Experiment completed
9 PIPELINES GENERATED
9 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 2 minutes
View log
Save code
Pipeline leaderboard
Rank
Name
Algorithm
Specialization
Accuracy (Optimized)
Cross Validation
Enhancements
Build time

RESULT



9 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

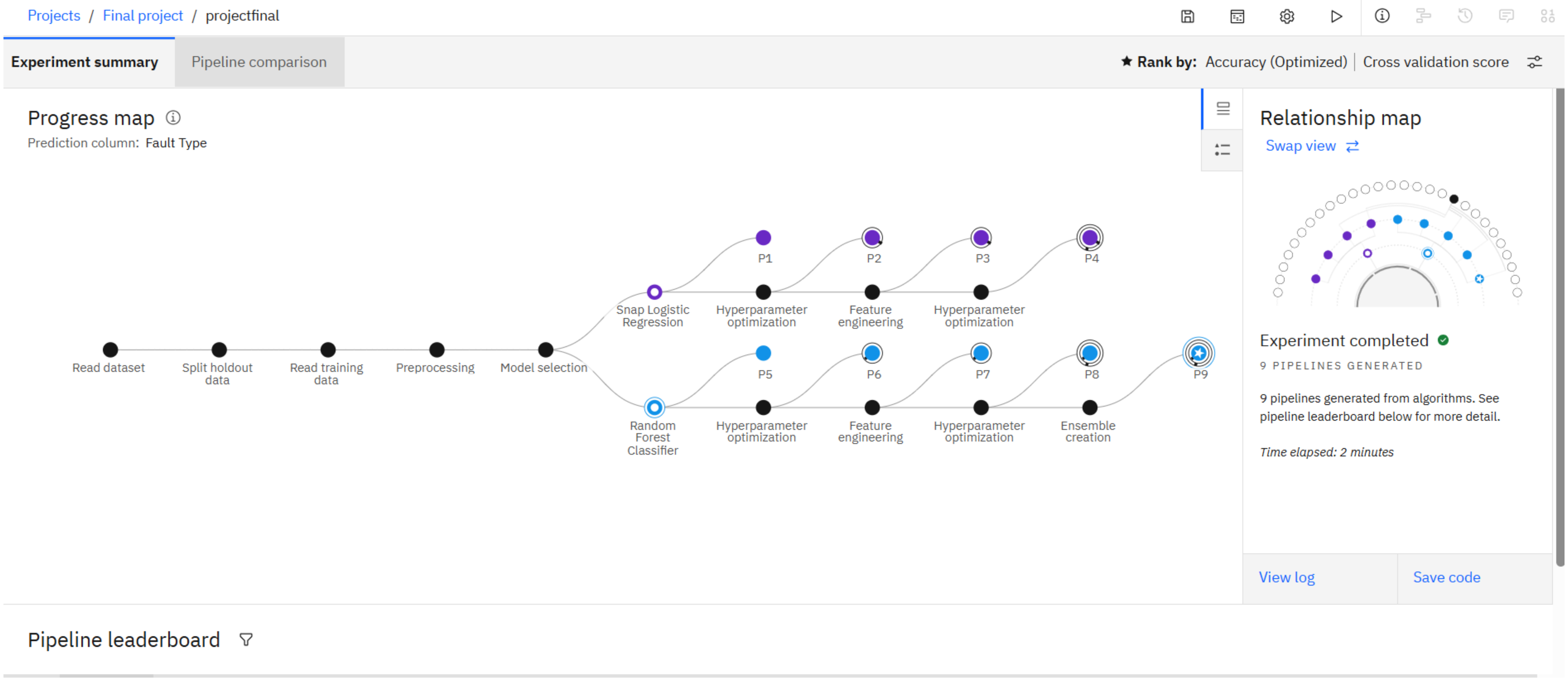
Time elapsed: 2 minutes

View log | Save code

Pipeline leaderboard

	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 9	Batched Tree Ensemble Classifier (Random Forest Classifier)	INCR	0.409	HPO-1 FE HPO-2 BATCH	00:00:48
	2	Pipeline 8	Random Forest Classifier		0.409	HPO-1 FE HPO-2	00:00:43
	3	Pipeline 4	Snap Logistic Regression		0.393	HPO-1 FE HPO-2	00:00:29
	4	Pipeline 3	Snap Logistic Regression		0.393	HPO-1 FE	00:00:24

CONCLUSION



Projects / Final project / projectfinal
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
Hyperparameter optimization
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: 2 minutes
View log
Save code
Pipeline leaderboard

CONCLUSION

9 PIPELINES GENERATED

9 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

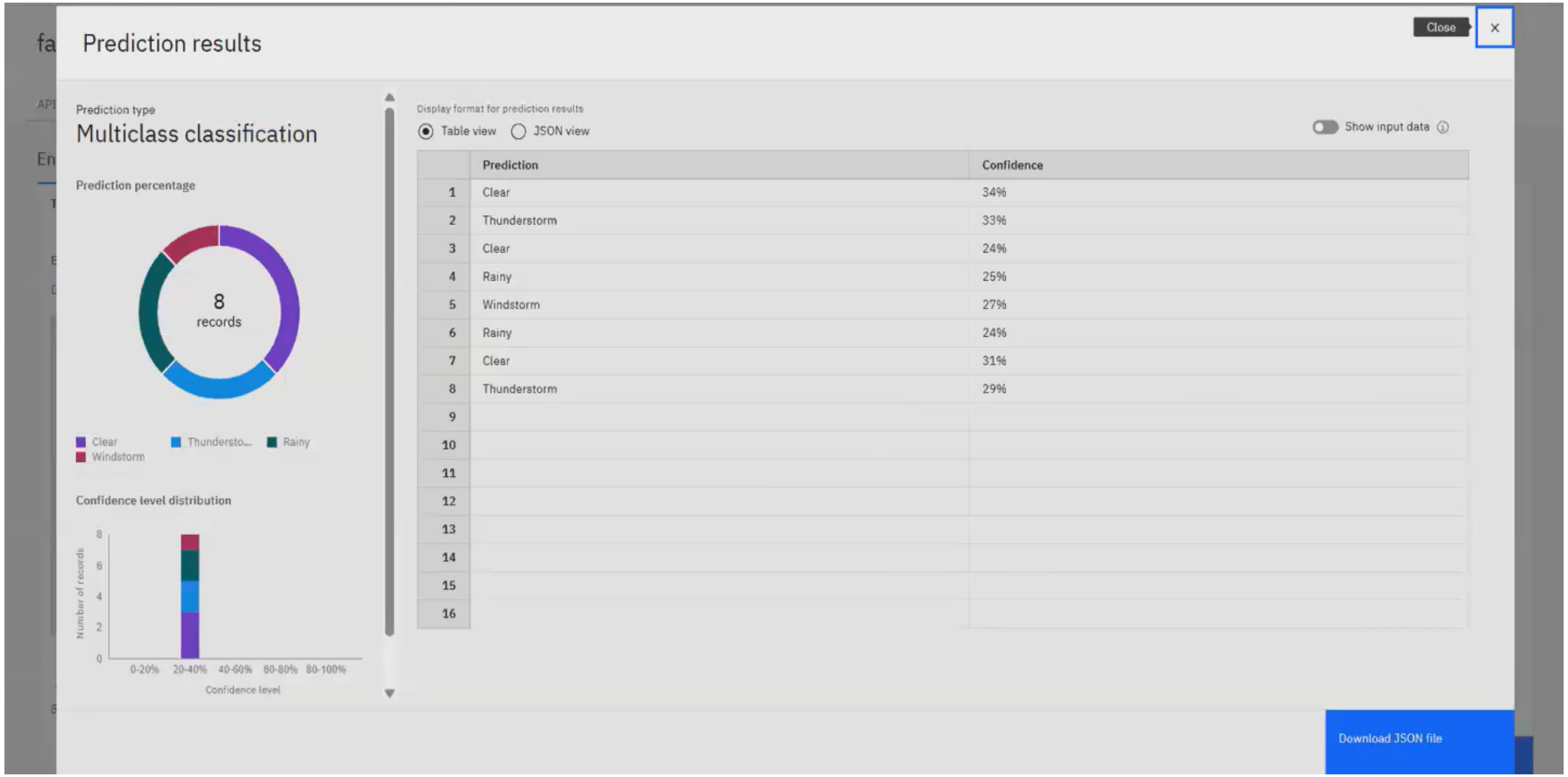
Time elapsed: 2 minutes

View log | Save code

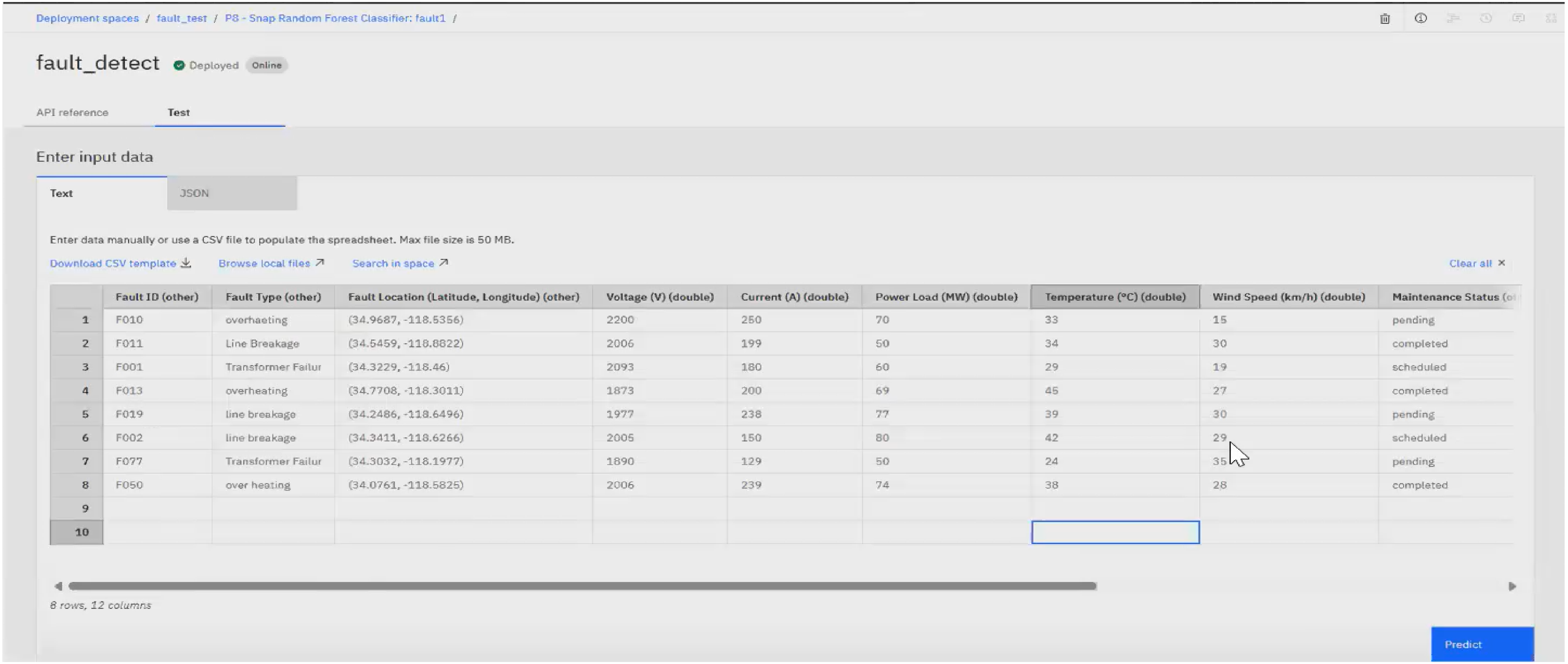
Pipeline leaderboard

	Rank ↑	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	**Pipeline 9**	Batched Tree Ensemble Classifier (Random Forest Classifier)	INCR	0.409	HPO-1 FE HPO-2 BATCH	00:00:48
	2	**Pipeline 8**	Random Forest Classifier		0.409	HPO-1 FE HPO-2	00:00:43
	3	**Pipeline 4**	Snap Logistic Regression		0.393	HPO-1 FE HPO-2	00:00:29
	4	**Pipeline 3**	Snap Logistic Regression		0.393	HPO-1 FE	00:00:24

FUTURE SCOPE



FUTURE SCOPE


Deployment spaces / fault_test / P8 - Snap Random Forest Classifier: fault1 /

fault_detect Deployed Online

API reference Test

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 Type (other)	Fault Location (Latitude, Longitude) (other)	Voltage (V) (double)	Current (A) (double)	Power Load (MW) (double)	Temperature (°C) (double)	Wind Speed (km/h) (double)	Maintenance Status (o
1	F010	overhaeting	(34.9687, -118.5356)	2200	250	70	33	15	pending
2	F011	Line Breakage	(34.5459, -118.8822)	2006	199	50	34	30	completed
3	F001	Transformer Failur	(34.3229, -118.46)	2093	180	60	29	19	scheduled
4	F013	overheating	(34.7708, -118.3011)	1873	200	69	45	27	completed
5	F019	line breakage	(34.2486, -118.6496)	1977	238	77	39	30	pending
6	F002	line breakage	(34.3411, -118.6266)	2005	150	80	42	29	scheduled
7	F077	Transformer Failur	(34.3032, -118.1977)	1890	129	50	24	35	pending
8	F050	over heating	(34.0761, -118.5825)	2006	239	74	38	28	completed
9									
10									

8 rows, 12 columns

Predict

edunet foundation

REFERENCES

- **J. Zhang et al.,** "A CNN-LSTM-based Fault Diagnosis Method for Power Transmission Lines," *IEEE Transactions on Power Systems*, 2021.
- **"Power System Fault Data" onggle** → Public datasets on platforms like Kaggle are invaluable for testing and validating different machine learning models. These datasets often contain simulated or real-world data of currents and voltages under various fault and normal conditions.

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

Sahaja Pallapothula

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/19be27f6-e1db-4a2d-ad69-79fcfdea63d6

IBM

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

Sahaja Pallapothula

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/a746141d-68cd-4a43-93b1-be49abf8d8c9

IBM

IBM CERTIFICATIONS


IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Sahaja Pallapothula

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