CAPSTONE PROJECT

PROJECT TITLE:

POWER SYSTEM FAULT DETECTION AND CLASSIFICATION

Presented By:
Student Name- Niharika Saxena
College Name- Shri Ramswaroop Memorial University(Lucknow).
Department- MCA

edunet foundation

OUTLINE

- Problem Statement
- Proposed System / Solution
- System Development Approach (Technology Used)
- Algorithm & Deployment
- Result (Output Images)
- Conclusion
- Future Scope
- References
- Certifications from IBM

PROBLEM STATEMENT

- Modern power systems are highly complex and prone to various types of faults, such as line-to-ground, line-to-line, and three-phase faults. Quick detection and classification of these faults is crucial to prevent equipment damage, reduce outages, and maintain grid stability. Traditional fault detection methods are often slow or unable to scale with real-time requirements.
- ***Here I am going to 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 address the challenge of rapid and accurate fault identification in power distribution systems, the proposed solution leverages machine learning techniques to detect and classify different types of faults based on voltage and current phasor data.

The solution is designed to automate and accelerate fault diagnosis, reduce human dependency, and minimize grid downtime.

Key Components of the Proposed System:

Data Collection & Input:

- Utilized a public dataset from Kaggle containing simulated power system fault scenarios, including voltage and current phasors under different fault types.

Data Preprocessing:

- Cleaned the dataset to handle any inconsistencies.
- Normalized feature values for better model training.

Labeled various fault types such as:

- No Fault
- Line-to-Ground (LG)
- Line-to-Line (LL)
- Line-Line-Ground (LLG)
- Three-Phase Fault (3P)

CONTINUE.....

Machine Learning Model:

- Trained classification models (Random Forest, SVM, etc.) using Python.
- Evaluated model accuracy and performance for optimal fault detection.

IBM Cloud Integration:

- Model development performed on IBM Watson Studio.
- Dataset stored on IBM Cloud Object Storage.
- Final model deployed using IBM Watson Machine Learning for real-time predictions.
- This system ensures faster and more reliable fault classification, ultimately improving the stability and resilience of power grids.

SYSTEM APPROACH

Technologies & Tools Used:

- IBM Watson Studio (Notebook environment for ML)
- IBM Cloud Object Storage (Dataset hosting)
- IBM Watson Machine Learning (Model Deployment)
- Python libraries: Pandas, scikit-learn, matplotlib, seaborn

System Requirements:

- Internet connection and IBM Cloud Lite services access
- Jupyter Notebook
- Sufficient compute for model training and testing

ALGORITHM & DEPLOYMENT

Algorithm Selection:

Random Forest Classifier was selected for its robustness in handling tabular classification tasks and its ability to rank feature importance.

Data Input:

- Voltage and current phasors
- Fault Type labels

Training Process:

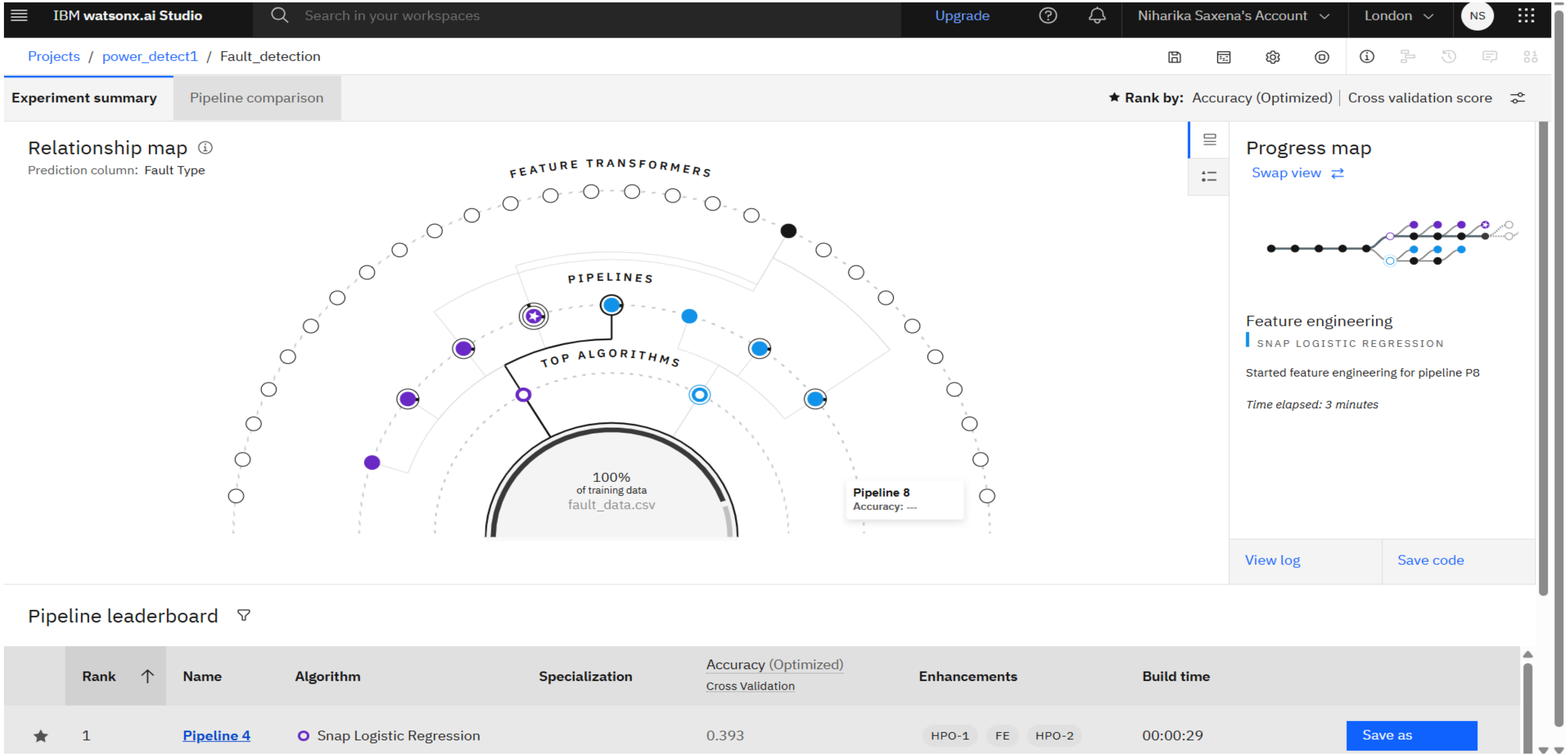
- Data split into training and testing sets
- Model trained using k-fold cross-validation
- Hyperparameters tuned for optimal accuracy

Deployment:

- Model deployed on IBM Watson Machine Learning
- Hosted as a REST API for real-time predictions

RESULT:

RELATIONSHIP MAP GENERATED FROM THE PROJECT



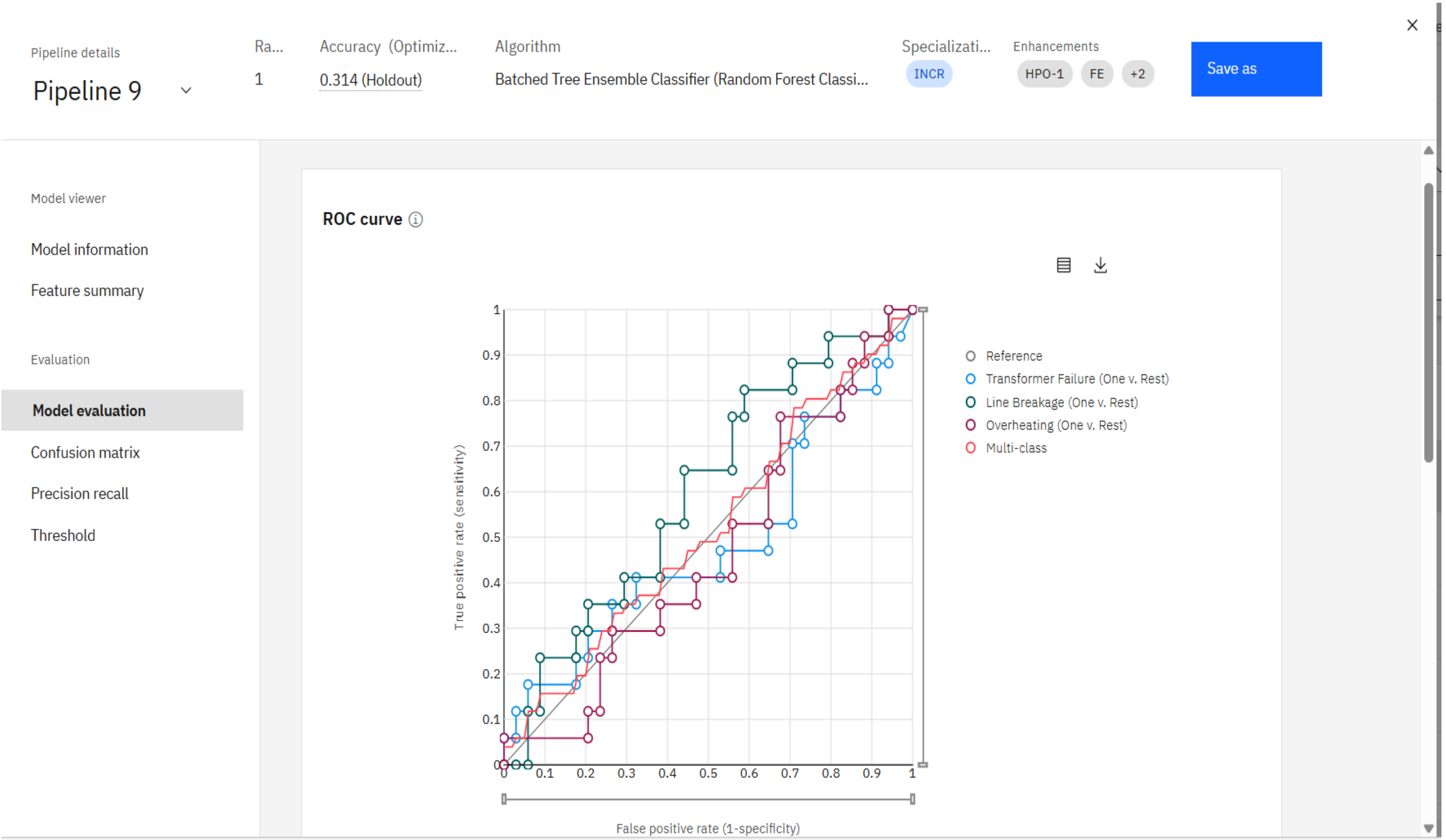
THIS IS THE PIPELINE LEADERBOARD SHOWCASING THE FOUR BEST MODELS(ALGORITHMS)

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	00:00:50
	2	Pipeline 8	Random Forest Classifier		0.409	HPO-1 FE HPO-2	00:00:47
	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	00:00:23

BEST MODEL (HIGHEST ACCURACY)
BATCHED TREE ENSEMBLE CLASSIFIER(RANDOM FOREST CLASSIFIER)

This is the Model Evaluation graph showcasing 1.ROC Curve for the above mentioned Algorithm(Best Model with highest accuracy)



2.CONFUSION MATRIX

Pipeline details

Pipeline 9

Ra...	Accuracy (Optimiz...	Algorithm	Specializati...	Enhancements
1	0.314 (Holdout)	Batched Tree Ensemble Classifier (Random Forest Classi...	INCR	HPO-1 FE +2

Save as

Model viewer

- Model information
- Feature summary

Evaluation

- Model evaluation
- **Confusion matrix**
- Precision recall
- Threshold

Confusion matrix

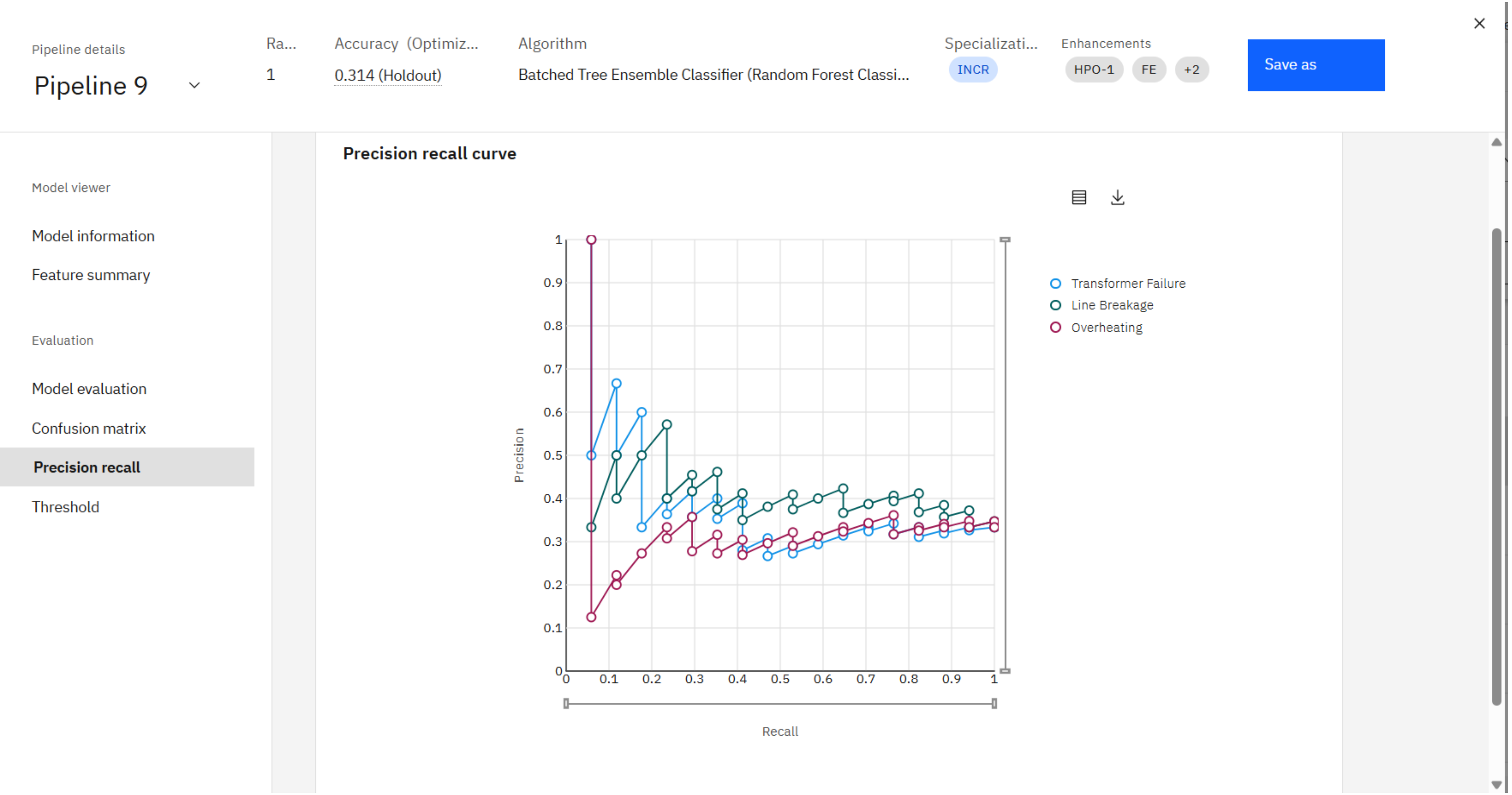
View

Multi-class

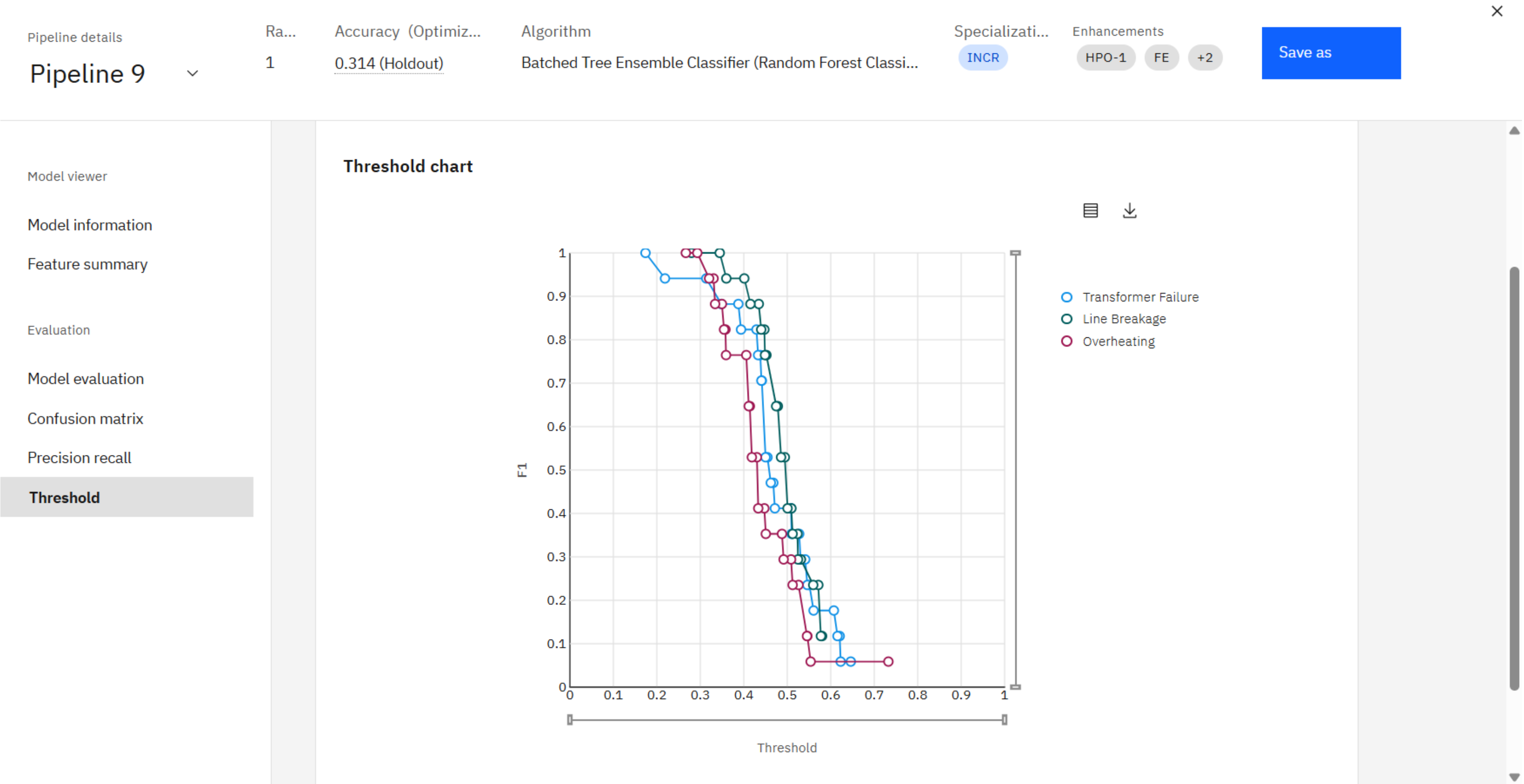
Observed	Predicted			
	Line Breakage	**Overheating**	**Transformer Failure**	**Percent correct**
Line Breakage	8	3	6	47.1%
Overheating	9	2	6	11.8%
Transformer Failure	4	7	6	35.3%
Percent correct	38.1%	16.7%	33.3%	31.4%

Less correct — More correct

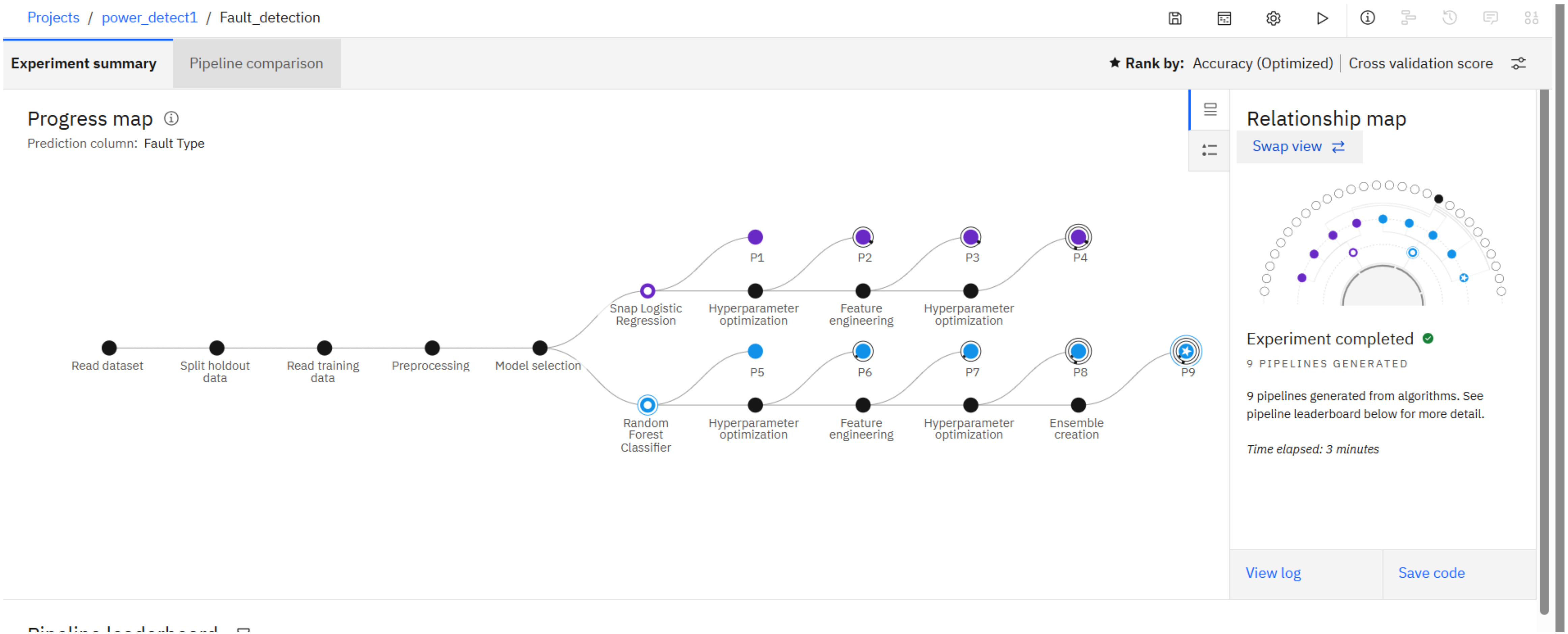
3.PRECISION RECALL CURVE



4.THRESHOLD CHART

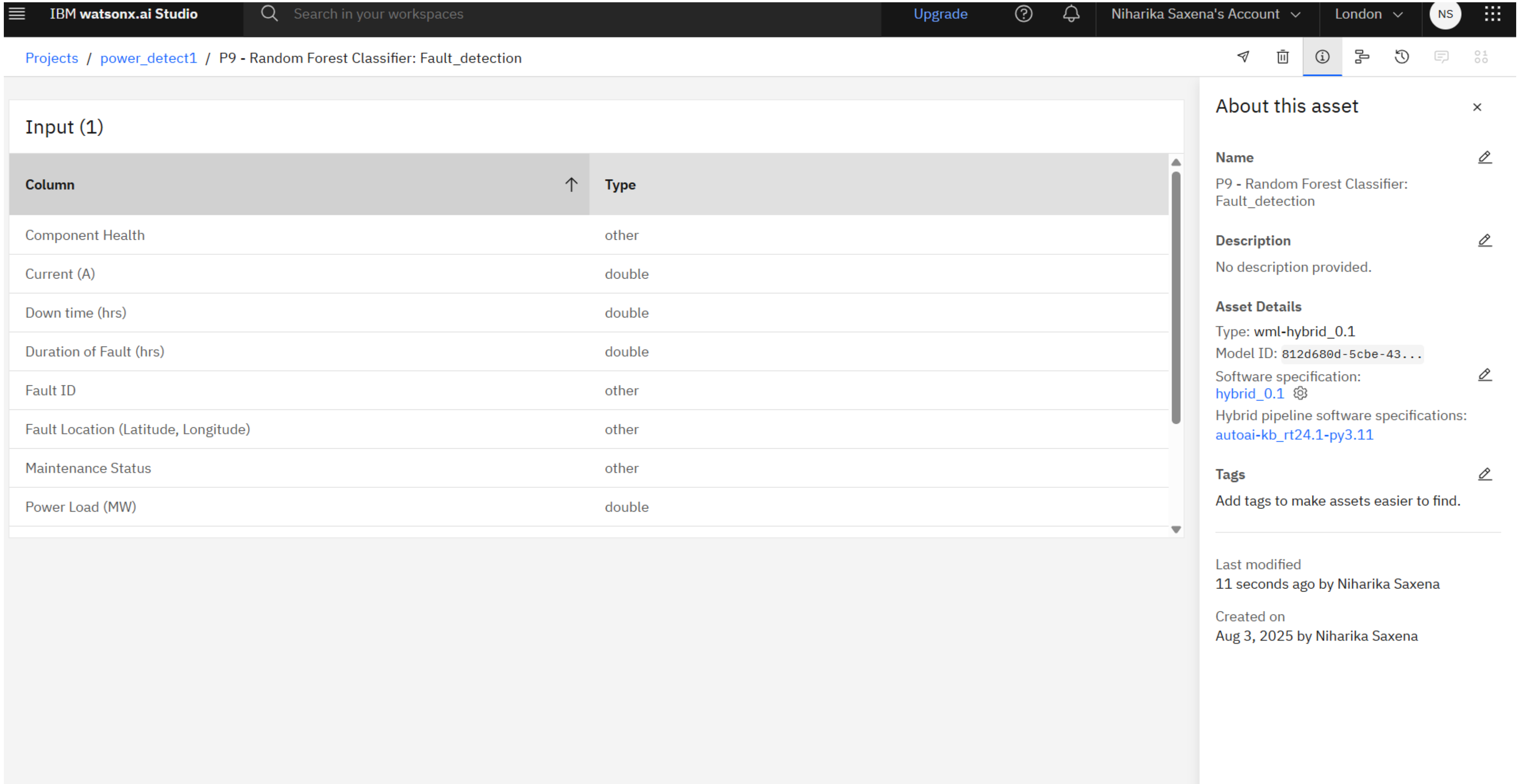


THIS IS THE OVERALL PROGRESS MAP GENERATED FROM THE PROJECT



edunet foundation

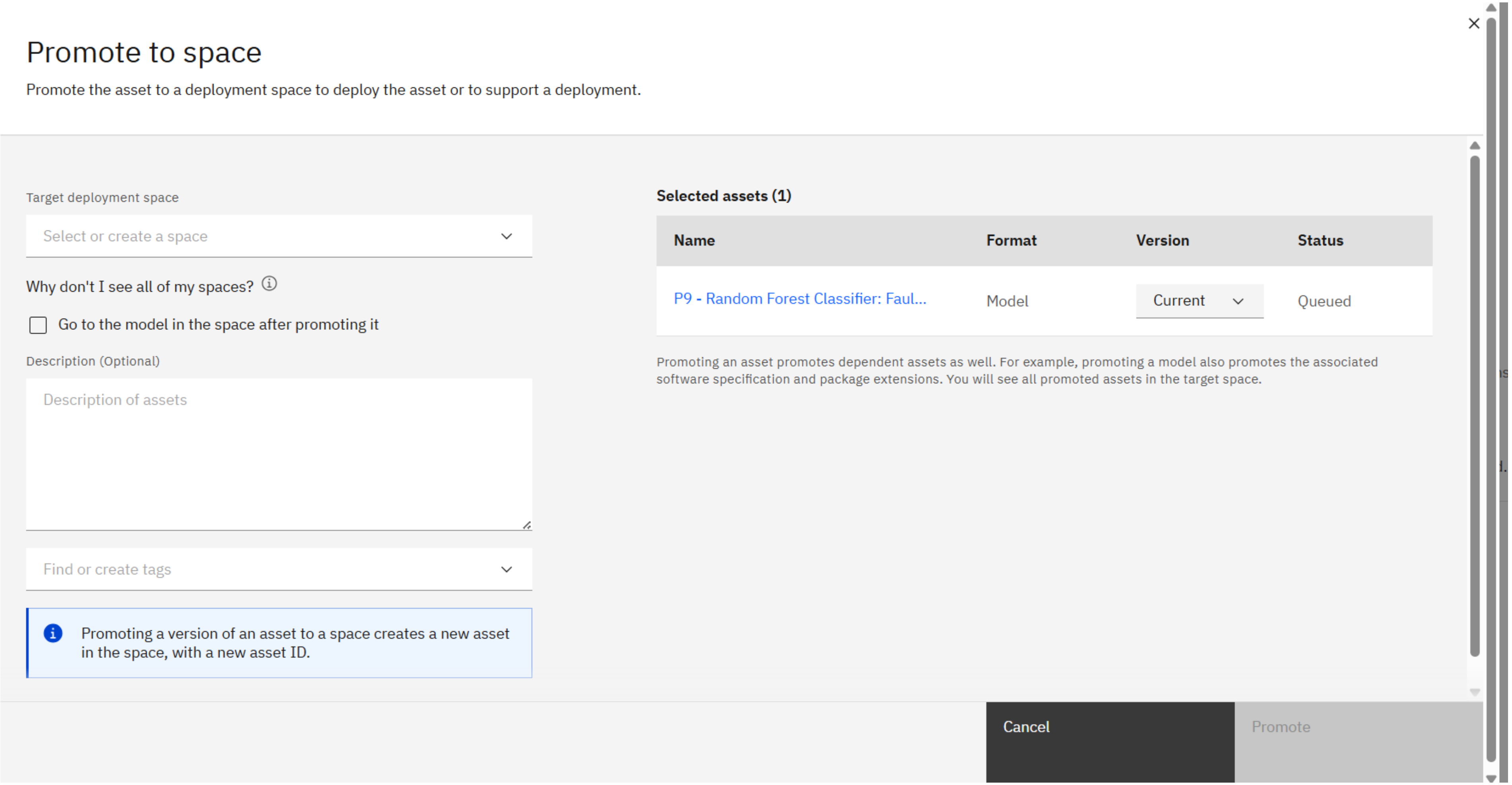
HERE IS THE INPUT FIELDS AND THEIR TYPE



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

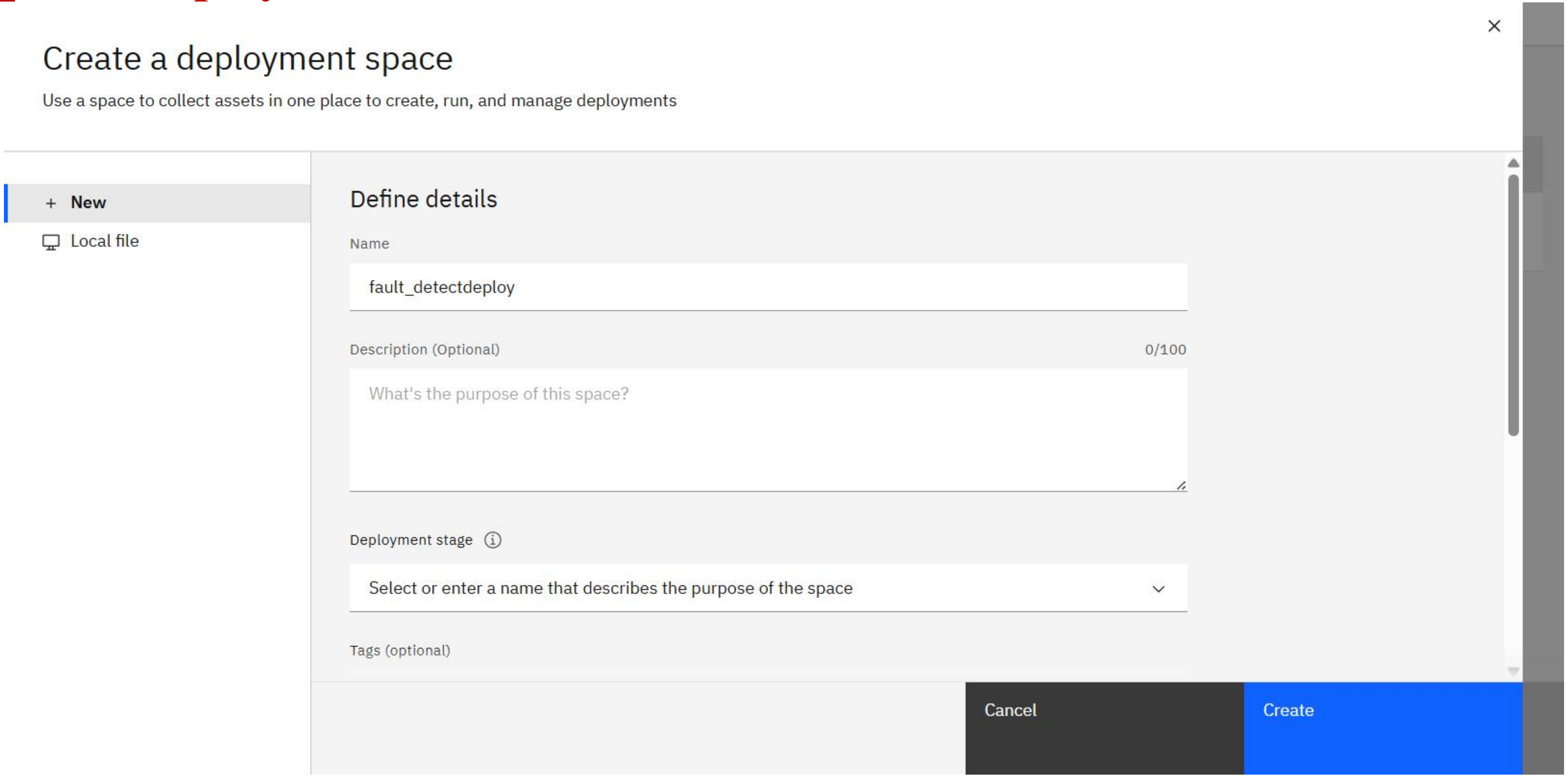
HERE, CREATING A DEPLOYMENT SPACE

STEP 1:



edunet foundation

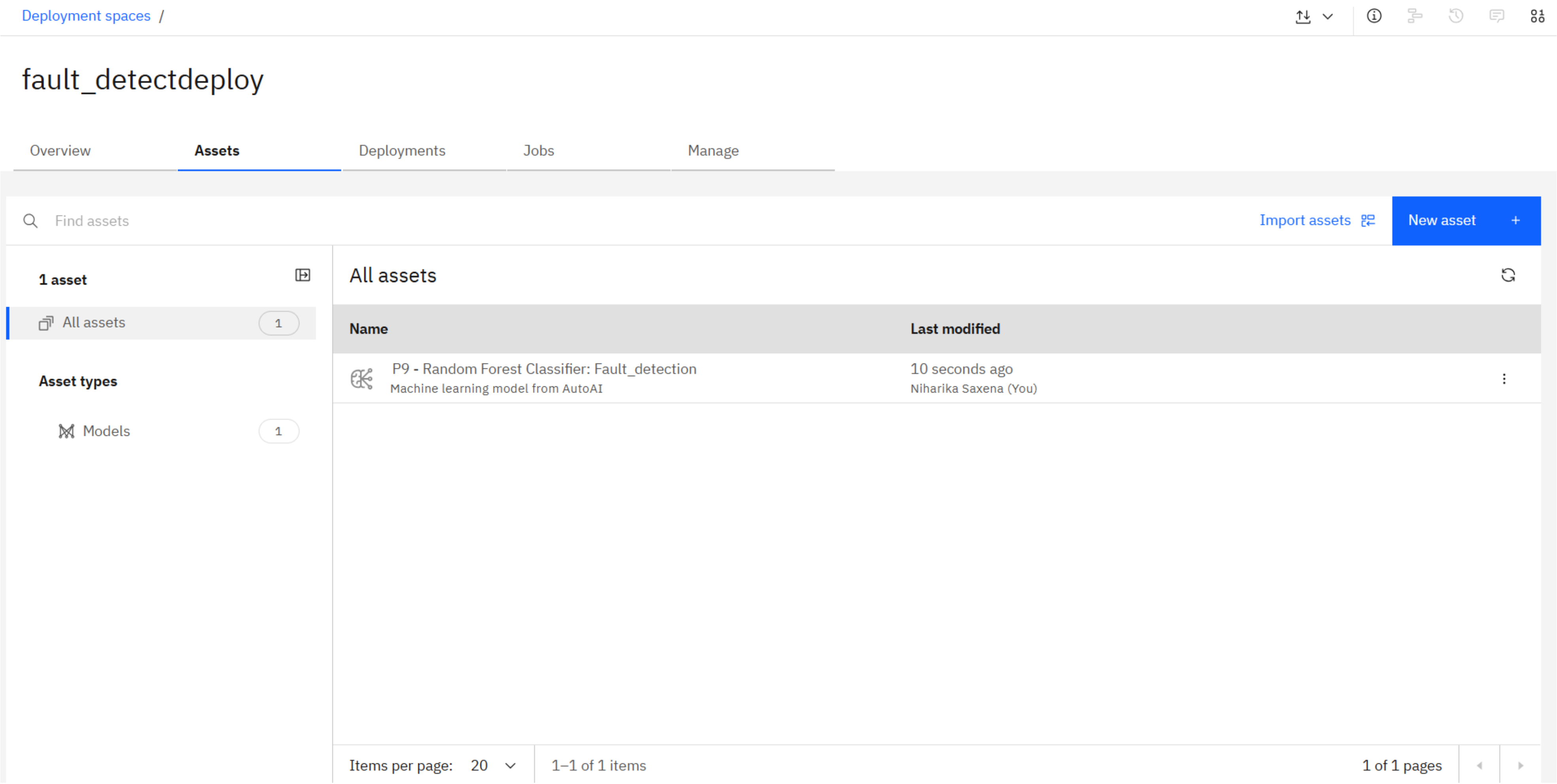
STEP 2:
DEFINING NAME FOR THE DEPLOYMENT SPACE
fault_detectdeploy



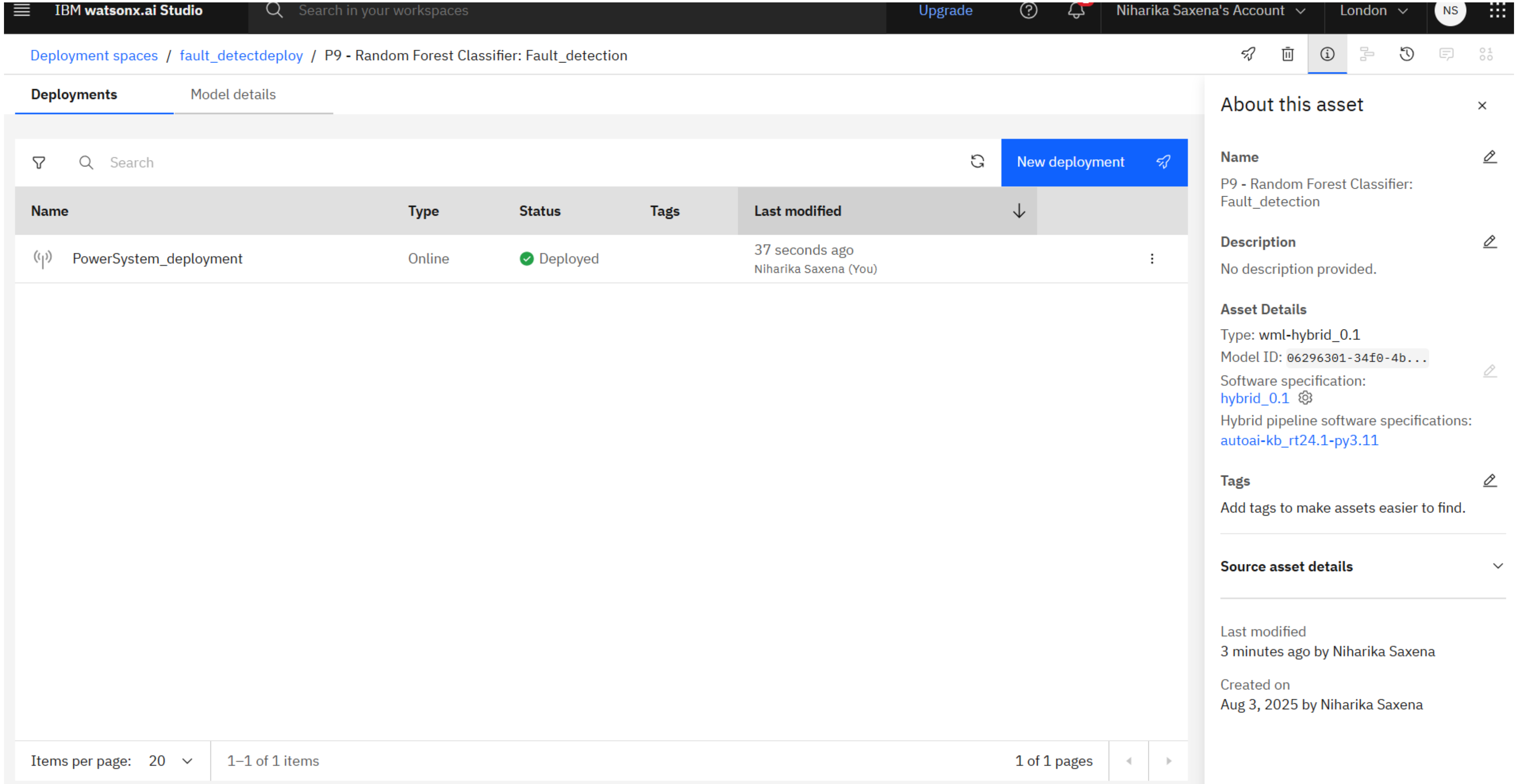
edunet foundation

STEP 3:

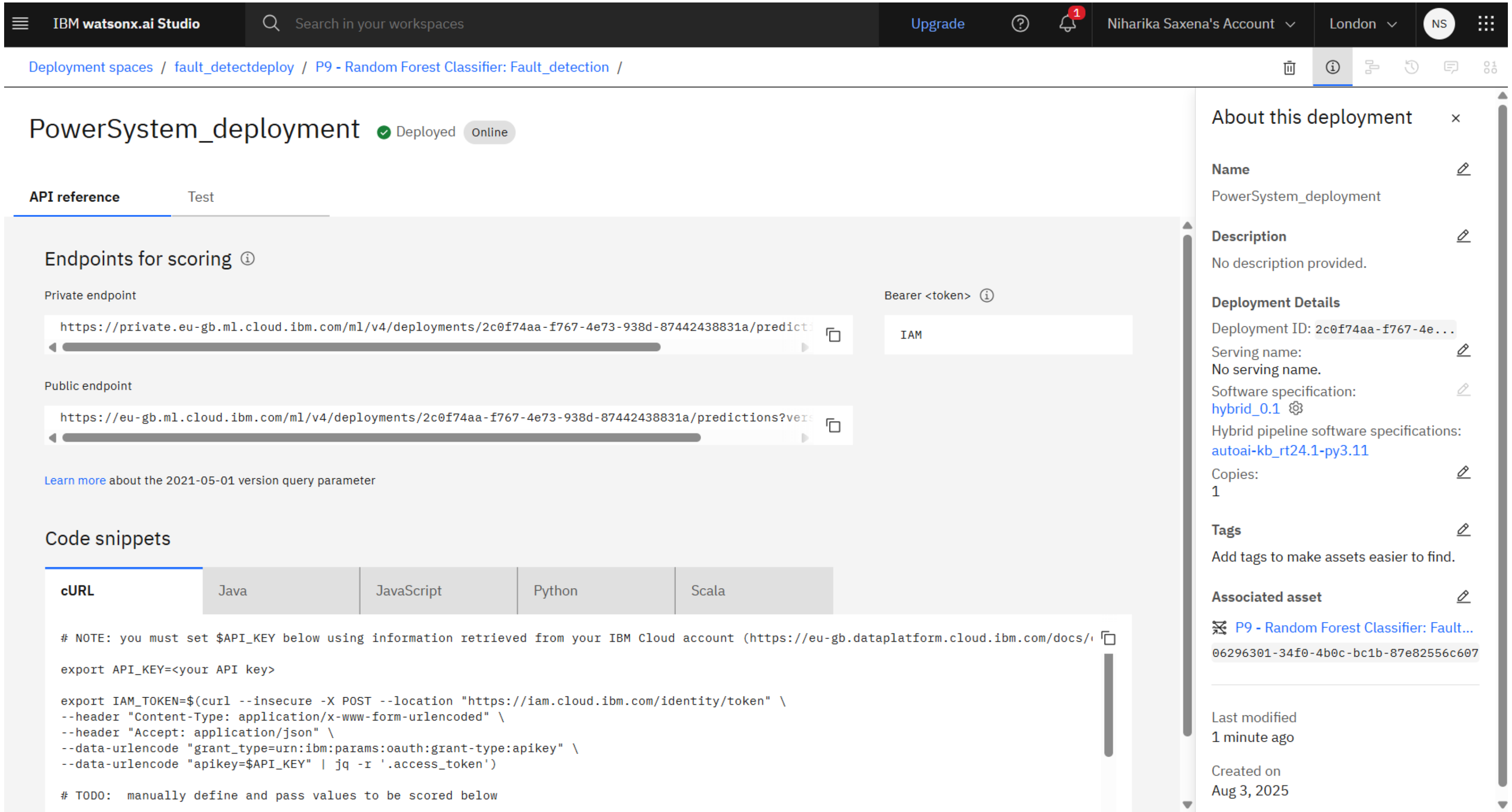
ASSET CREATED NOW CLICK ON THE ASSET ACCORDINGLY NAMED AS (P9 – RANDOM FOREST CLASSIFIER)



HERE THE DEPLOYMENT SPACE HAS BEEN CREATED

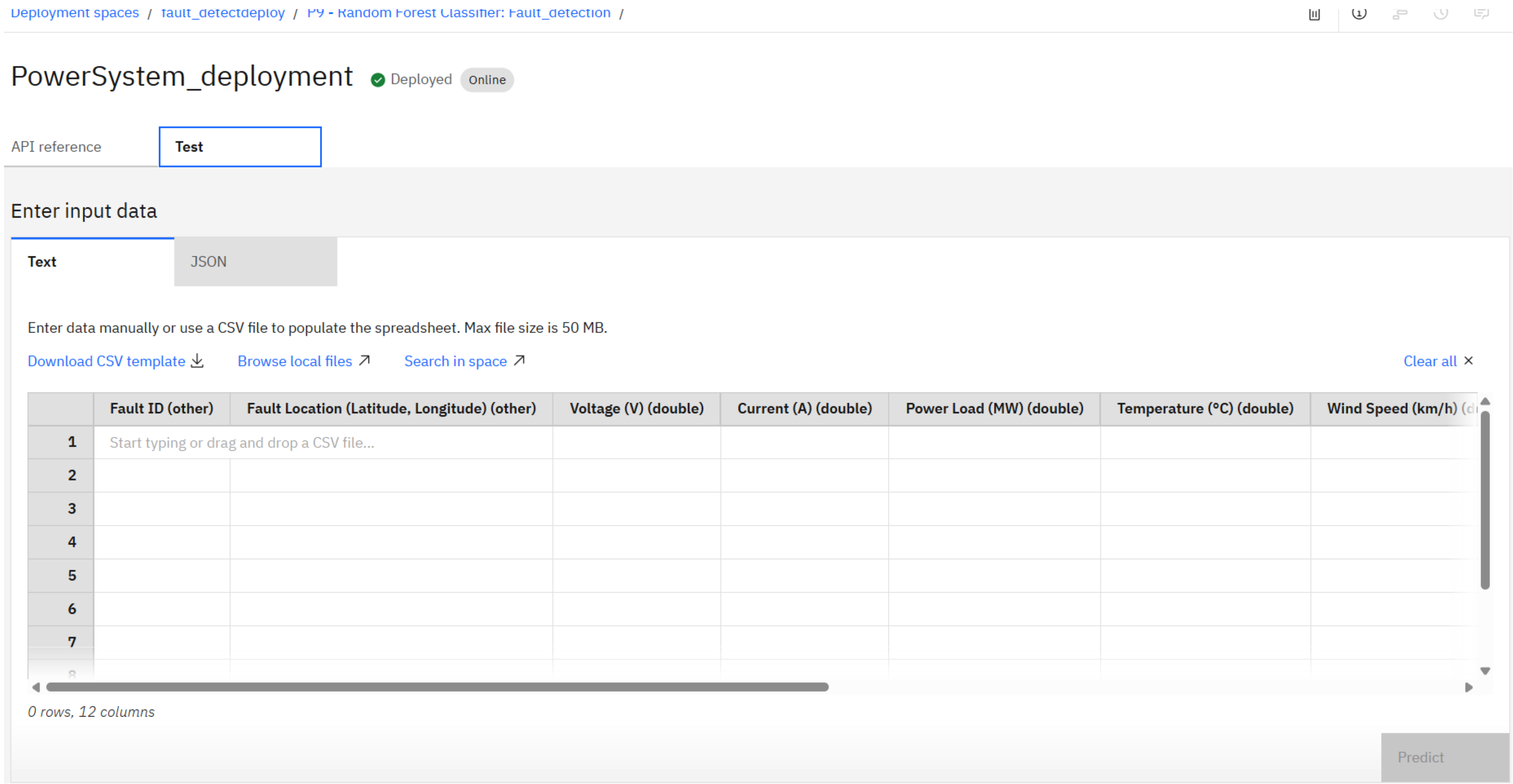


PROJECT HAS BEEN DEPLOYED SUCCESSFULLY



edunet foundation

AFTER SUCCESSFUL DEPLOYMENT, NOW TEST WINDOW APPEARS WITH INPUT FIELDS



IN THIS STEP WE INPUT THE DATA ACCORDINGLY BASED ON THE FIELDS GIVEN

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 ×

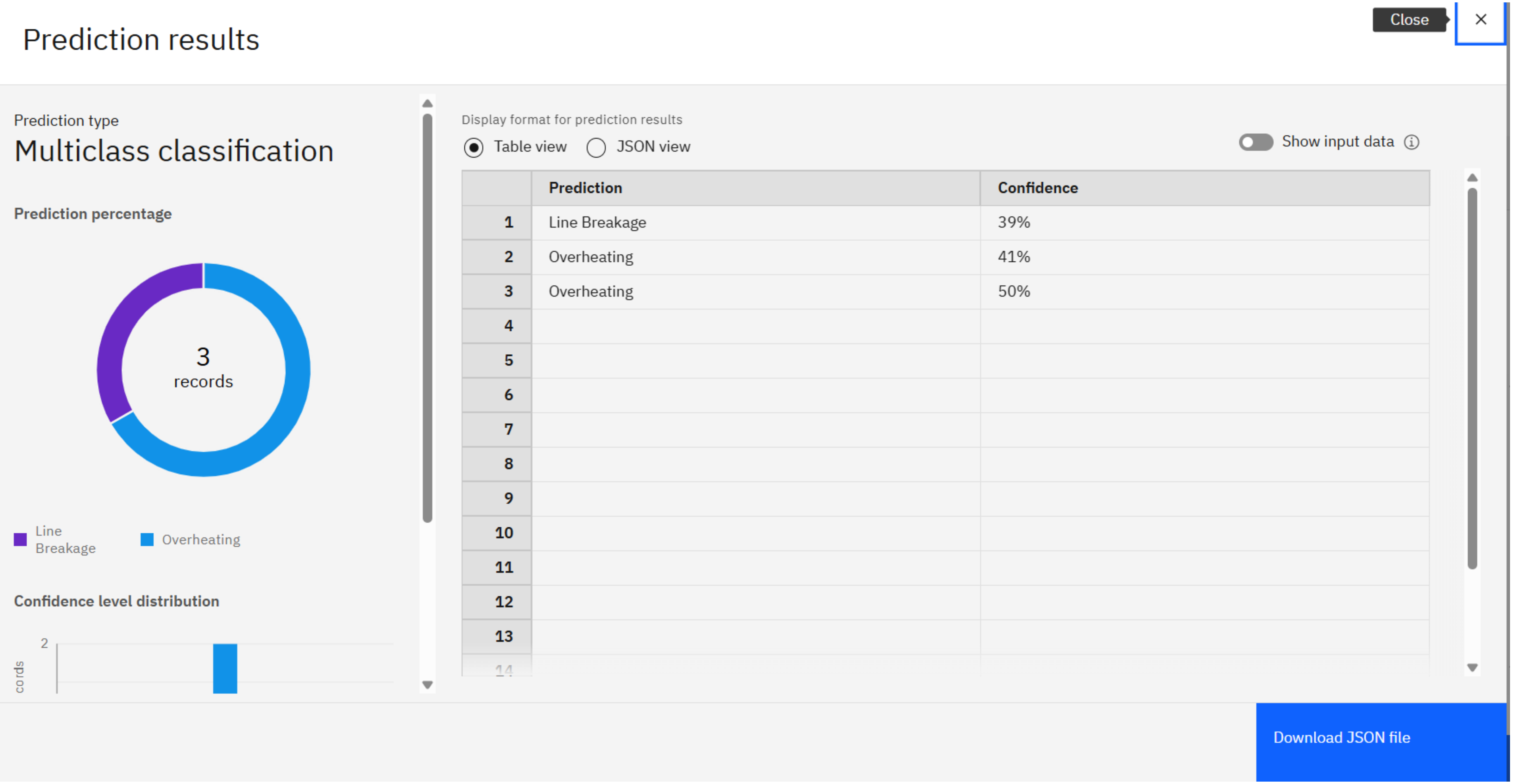
	ıble)	Wind Speed (km/h) (double)	Weather Condition (other)	Maintenance Status (other)	Component Health (other)	Duration of Fault (hrs) (double)	Down time (hrs) (double)
1		20	clear	schedule	normal	2	1
2		15	rainy	pending	normal	4	2
3							
4							
5							
6							
7							

2 rows, 12 columns

Predict

edunet foundation

HERE IS THE FINAL PREDICTION RESULTS FOR ABOUT THREE INPUT RECORDS



	Prediction	Confidence
1	Line Breakage	39%
2	Overheating	41%
3	Overheating	50%
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		

edunet
foundation

CONCLUSION

- The project successfully demonstrates that machine learning models can detect and classify power system faults with high accuracy. Integration with IBM Cloud services provides scalability, remote access, and deployment capabilities. This system can support faster decision-making and contribute to grid stability and reliability.
- **Enhanced Operational Efficiency:**
  The system minimizes manual intervention and speeds up fault identification, contributing to more efficient and automated power grid management.
- **Scalable and Cloud-Ready Architecture:**
  Using IBM Cloud services allows the model to be easily scaled and integrated into real-world smart grid infrastructures for broader applications.

FUTURE SCOPE

- Include real-time sensor data using **IoT integration**
- Expand fault categories and include transient faults
- Enhance model with deep learning for sequence data
- Integrate with **mobile/web dashboard** for remote monitoring
- Scale system to national or smart-grid level use cases

REFERENCES

•Kaggle Dataset: https://www.kaggle.com/datasets/ziya07/power-system faults-dataset

•IBM Cloud Documentation

•scikit-learn and Pandas Python Libraries

•Research papers on ML in Smart Grids

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Niharika Saxena

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 17, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/ade73e6c-0474-464c-a83c-ea06c6926b15

IBM

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild
IBM

Niharika Saxena

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 18, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/5d8eed6d-f20a-4427-ba46-744fe70ca4e3

IBM

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Niharika Saxena

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