CAPSTONE PROJECT

POWER SYSTEM FAULT DETECTION AND CLASSIFICATION

Presented By:
1. TANISHA CHOPRA – SG. BALEKUNDRI INSTITUTE OF TECHNOLOGY

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- 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

The system will use machine learning to detect and classify power system faults. It will help identify faults like Line-to-Ground, Line-to-Line, and Three-Phase quickly and accurately.

- **1. Data Collection:** Historical data of voltage and current phasors will be used. The Kaggle fault dataset will be the main source.
- **2. Data Preprocessing:** The data will be cleaned, normalized, and split into training and testing sets. Important features like voltage and current will be selected.
- **3. Machine Learning Model:** A classification model such as Random Forest or SVM will be trained to predict fault types. The model will classify connections as Normal or Faulty (with specific fault types).
- **4. Deployment:** A simple dashboard or app will be created to input values and display fault predictions. The system will be deployed on IBM Cloud.
- **5. Evaluation:** The model will be evaluated using accuracy, precision, recall, and a confusion matrix. Results will show how well the model classifies each fault type. Let me know if you want this added to your report or abstract too!

SYSTEM APPROACH

System Requirements:

- **Hardware Requirements:**
  - Minimum 4 GB RAM (locally)
  - Stable internet connection (for IBM Cloud use)

- **IBM Cloud Requirements:**
  - IBM Cloud Lite account (Free tier)
  - Watson Studio instance
  - Cloud Object Storage (for storing datasets)

ALGORITHM & DEPLOYMENT

Algorithm

- **Algorithm Selection:**

Random Forest is used for fault classification due to its accuracy, robustness, and ability to handle complex, multiclass data.

- **Data Input:**

The model uses 41 features including voltage, current, angle, and phasor values. The target is the fault_type, such as:

Line-to-Ground

Line-to-Line

Three-Phase

Normal

- **Training Process:**

Data is cleaned, normalized, and split (80/20). The model is trained using cross-validation and hyperparameter tuning to improve accuracy.

- **Prediction Process:**

Real-time input data is passed into the trained model, which predicts the fault type with a confidence score.

ALGORITHM & DEPLOYMENT

Deployment

- **Platform:**

The model is deployed using IBM Cloud services:

Watson Studio (training)

Cloud Object Storage (data)

Watson Machine Learning (deployment)

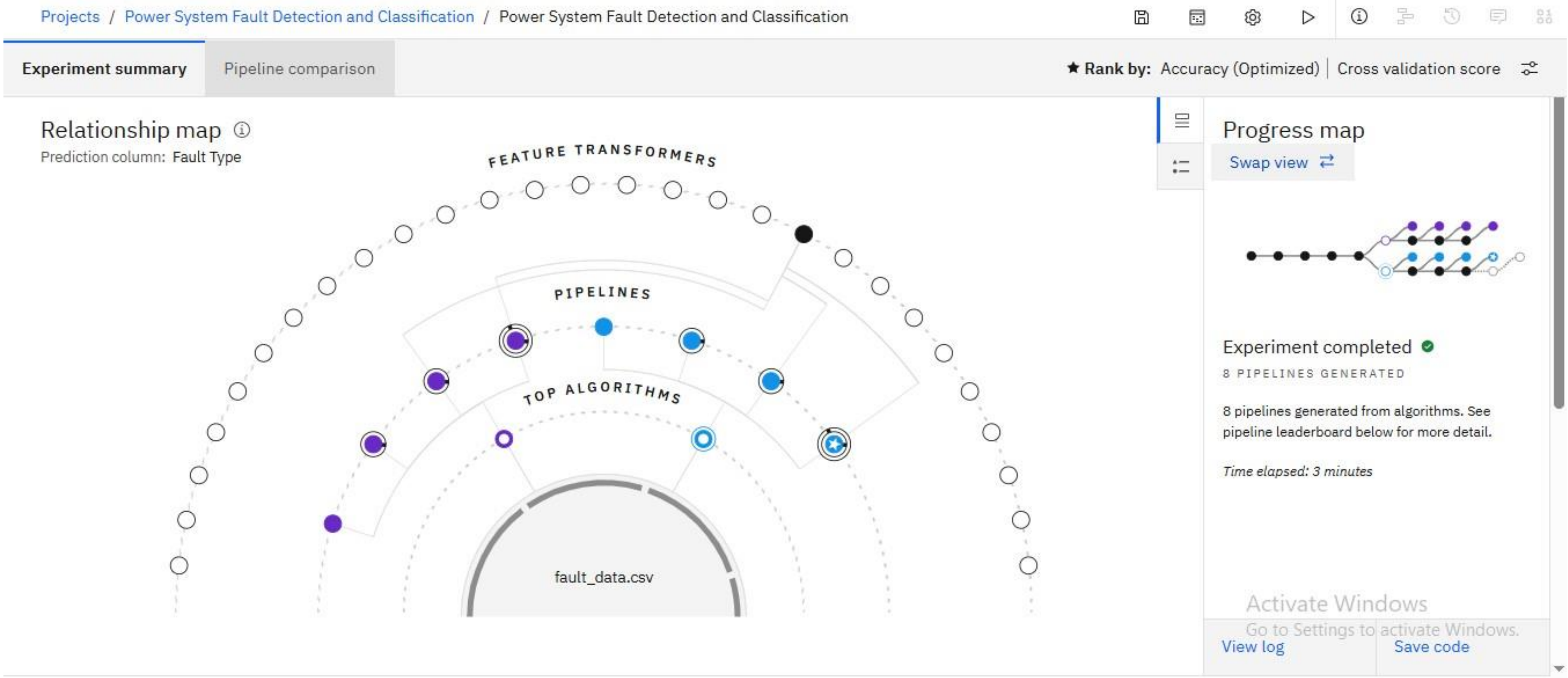
- **Steps:**

Upload and train the model in Watson Studio.

Deploy it as a REST API using Watson Machine Learning.

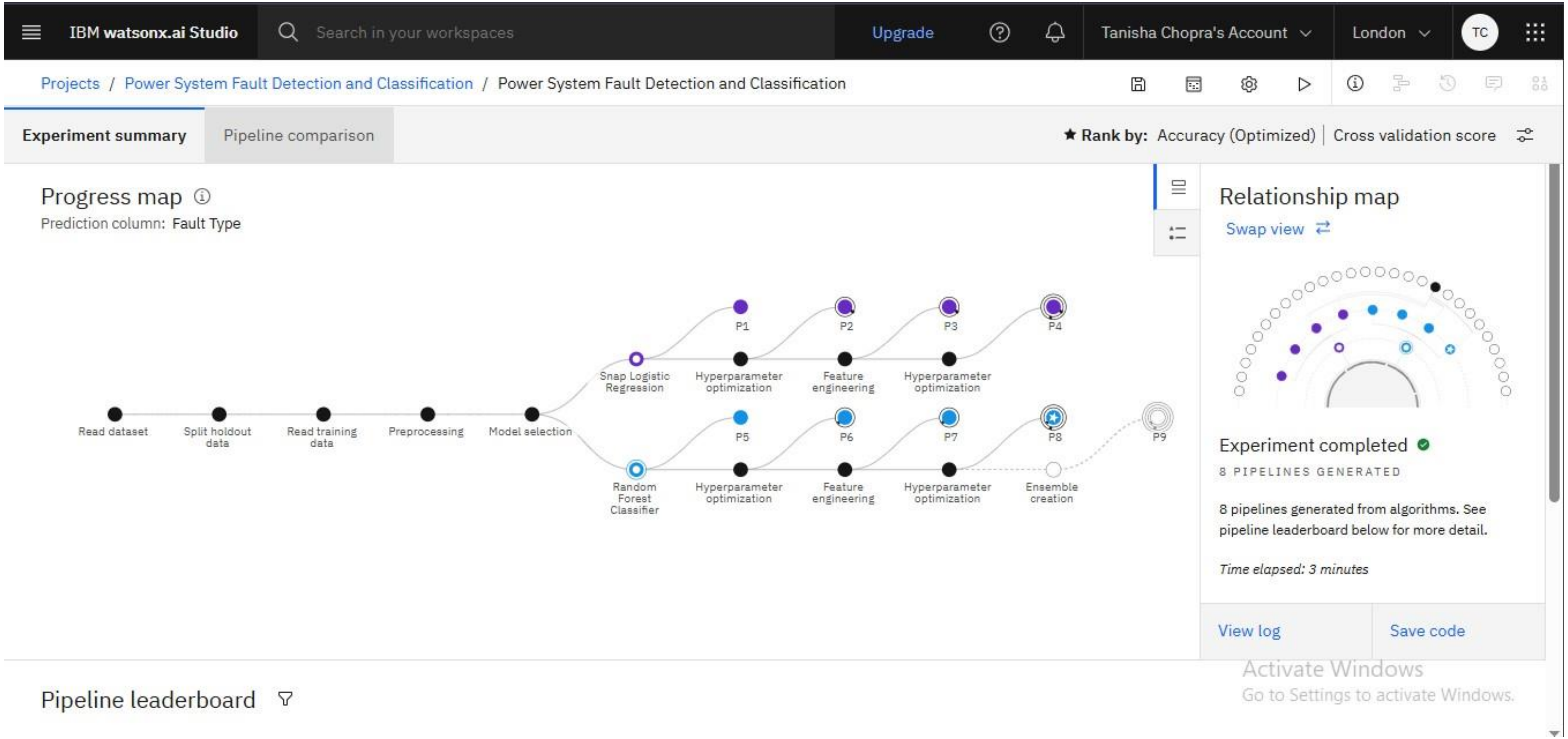
Use test data or a web app to send inputs and receive fault predictions.

RESULT



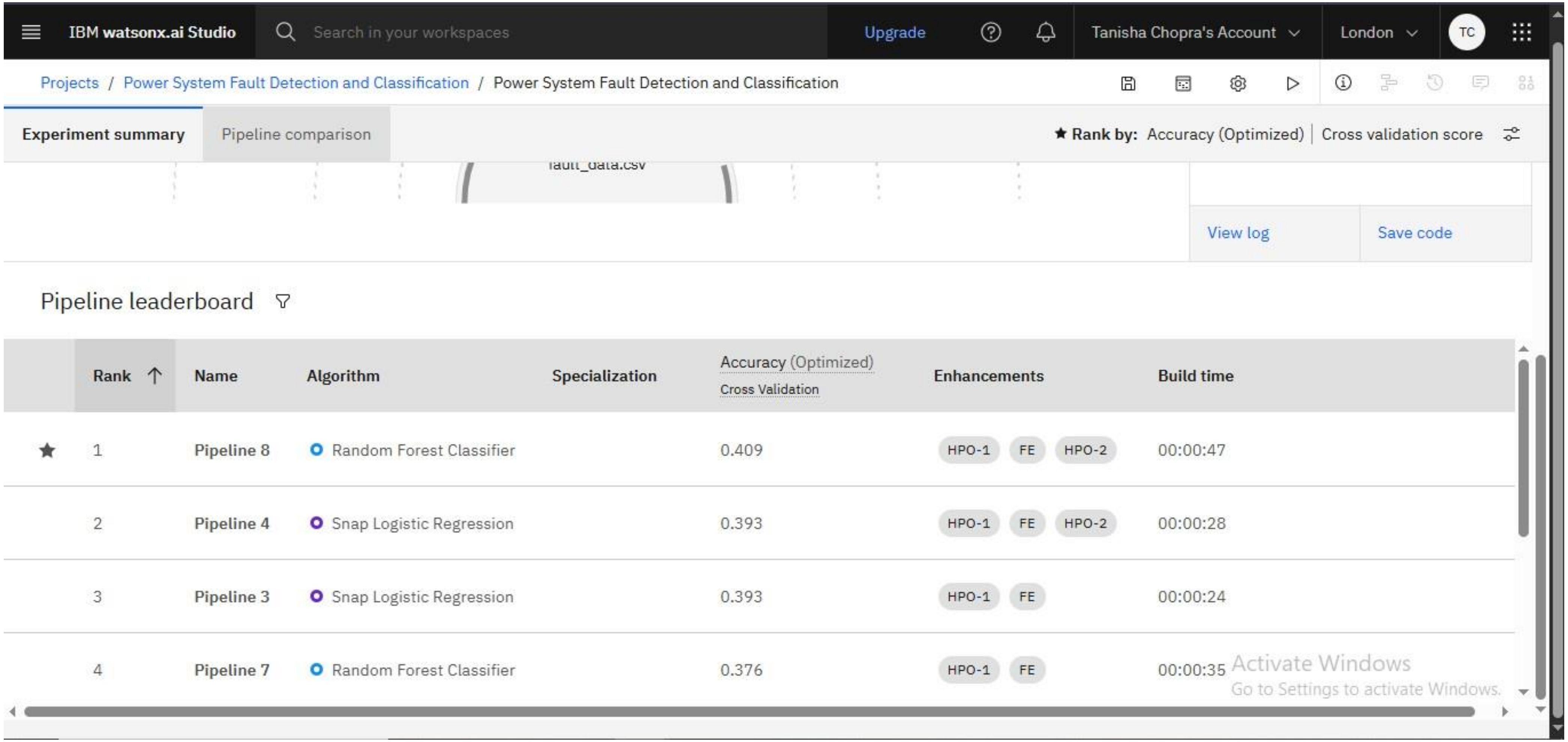
edunet
foundation

RESULT

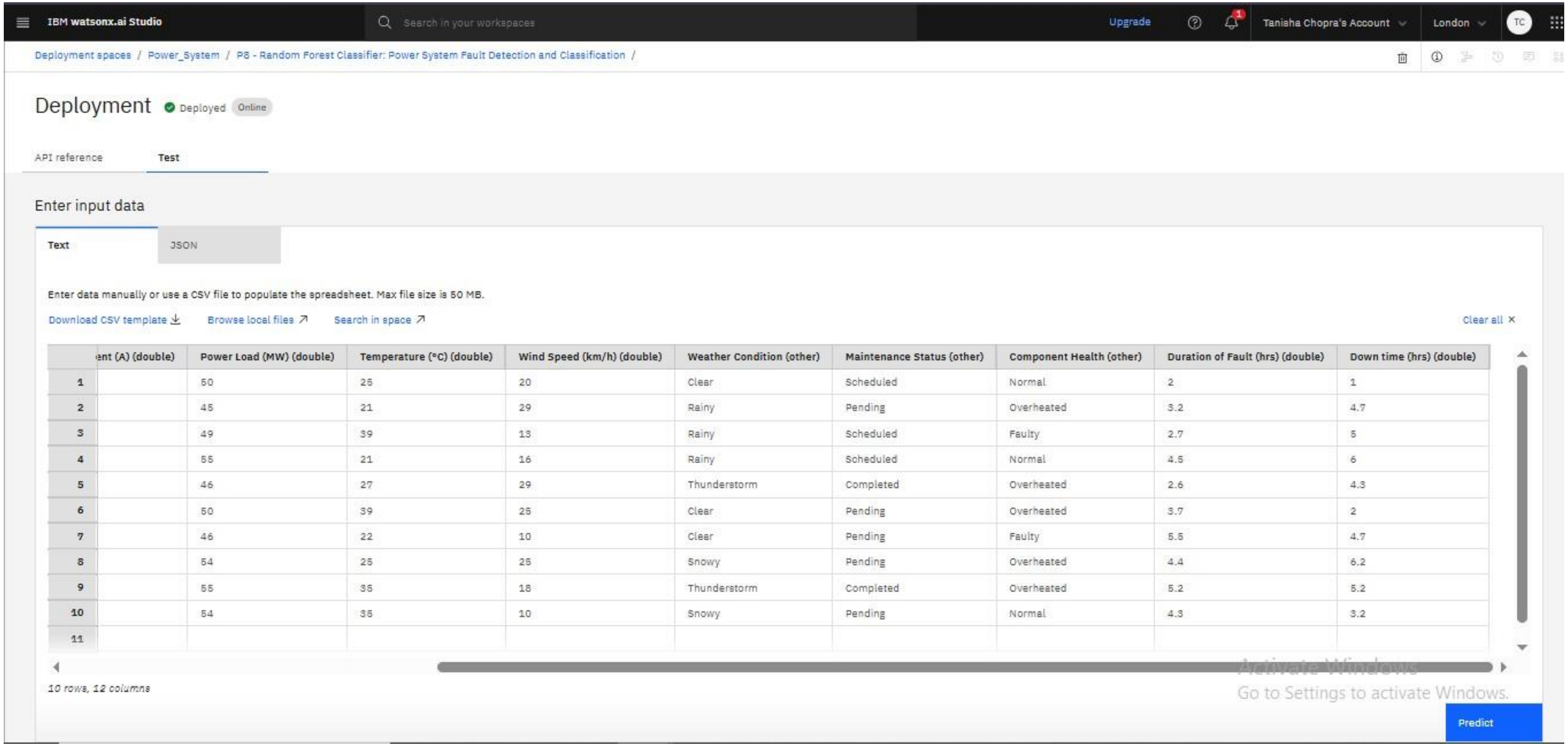


edunet foundation

RESULT


IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Tanisha Chopra's Account
London
TC
Projects / Power System Fault Detection and Classification / Power System Fault Detection and Classification
Experiment summary
Pipeline comparison
★ Rank by: Accuracy (Optimized) | Cross validation score
View log
Save code
Pipeline leaderboard
Rank
Name
Algorithm
Specialization
Accuracy (Optimized)
Cross Validation
Enhancements
Build time
1 Pipeline 8 Random Forest Classifier 0.409 HPO-1 FE HPO-2 00:00:47
2 Pipeline 4 Snap Logistic Regression 0.393 HPO-1 FE HPO-2 00:00:28
3 Pipeline 3 Snap Logistic Regression 0.393 HPO-1 FE 00:00:24
4 Pipeline 7 Random Forest Classifier 0.376 HPO-1 FE 00:00:35
Activate Windows
Go to Settings to activate Windows.

RESULT



	ent (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		50	25	20	Clear	Scheduled	Normal	2	1
2		45	21	29	Rainy	Pending	Overheated	3.2	4.7
3		49	39	13	Rainy	Scheduled	Faulty	2.7	5
4		55	21	16	Rainy	Scheduled	Normal	4.5	6
5		46	27	29	Thunderstorm	Completed	Overheated	2.6	4.3
6		50	39	25	Clear	Pending	Overheated	3.7	2
7		46	22	10	Clear	Pending	Faulty	5.5	4.7
8		54	25	25	Snowy	Pending	Overheated	4.4	6.2
9		55	35	18	Thunderstorm	Completed	Overheated	5.2	5.2
10		54	35	10	Snowy	Pending	Normal	4.3	3.2
11									

edunet foundation

RESULT

Prediction results

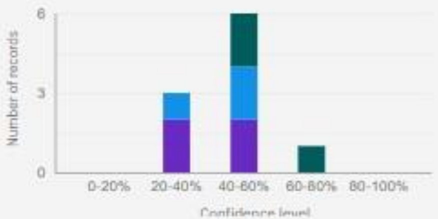
×

Prediction type

Multiclass classification

Prediction percentage

Confidence level distribution



Display format for prediction results

Table view   JSON view

Show input data

	Prediction	Confidence
1	Line Breakage	39%
2	Transformer Failure	41%
3	Line Breakage	42%
4	Overheating	41%
5	Transformer Failure	38%
6	Overheating	48%
7	Overheating	63%
8	Line Breakage	40%
9	Transformer Failure	43%
10	Line Breakage	44%
11		
12		
13		
14		
15		
16		

Download JSON file

CONCLUSION

- The proposed machine learning-based system successfully detects and classifies different types of power system faults using real-time electrical data. By leveraging algorithms like Random Forest and deploying the model on IBM Cloud, the system ensures fast and accurate fault identification, which is essential for maintaining power grid stability and reducing downtime. The cloud-based deployment also supports scalability, accessibility, and integration with smart grid infrastructure.

FUTURE SCOPE

- Integrate the system with **real-time IoT sensors** in smart grids.
- Extend the model to **predict fault severity and recovery time.**
- Include **deep learning** methods like LSTM for better fault prediction in dynamic environments.
- Build a **mobile/web dashboard** for real-time monitoring and alerts.
- Improve the model with **more diverse datasets** from various regions or seasons.

REFERENCES

- Kaggle – *Power System Faults Dataset*
  https://www.kaggle.com/datasets/ziya07/power-system-faults-dataset
- IBM Cloud Docs – *Watson Studio*
  https://www.ibm.com/cloud/watson-studio
- IBM Cloud - Watson Studio
  https://www.ibm.com/cloud/watson-studio
- Scikit-learn Documentation – *Random Forest Classifier*
  https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
- IBM Cloud - Watson Machine Learning
  https://www.ibm.com/cloud/machine-learning

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

TANISHA DHIRAJ CHOPRA

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/efa7b4bf-3139-4bde-9cdf-4cd3d5528af4

IBM

edunet foundation

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM

TANISHA DHIRAJ CHOPRA

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/9c7422d1-b274-444e-8658-2a94d97cea67

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

TANISHA DHIRAJ CHOPRA

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT)

Learning hours: 20 mins

THANK YOU