CAPSTONE PROJECT

POWER SYSTEM FAULT DETECTION AND CLASSIFICATION

Presented By:
1. Arpan Tiwari-Ajay Kumar Garg Engineering College-CSE

OUTLINE

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

The operational integrity of modern power distribution systems is critically dependent on the rapid and accurate detection of electrical faults. Conventional protection schemes, often face challenges in speed, selectivity, and adaptability to the dynamic nature of grid operations. Delays or misclassifications of fault events—such as Line-to-Ground (LG), Line-to-Line (LL), Double Line-to-Ground (LLG), or Three-Phase (LLL) faults—can lead to extended power outages, equipment damage, and potential grid instability. Therefore, there is a pressing need to develop an intelligent, data-driven system capable of autonomously analysing system parameters to provide instantaneous and precise fault classification.

PROPOSED SOLUTION

- This project aims to design, train, and validate a robust machine learning model that utilizes synchronized electrical measurement data, specifically voltage and current phasors, to automatically distinguish between normal operating conditions and various fault typologies. The solution will consist of the following components:
- Data Collection:
  - Use the Kaggle dataset on power system faults(.csv file).
- Data Preprocessing:
  - Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.
- Machine Learning Algorithm:
  - Implement the data and train a classification model (e.g., Decision Tree, Random Forest or SVM).
  - Consider incorporating other factors like weather conditions, wind speed, and temperature to improve prediction accuracy.
- Deployment:
  - Deploy the solution on a scalable and reliable platform (IBM Cloud).
- Evaluation:
  - Assess the model's performance using appropriate metrics such as accuracy, precision, recall and F1-score.
  - Fine-tune the model based on feedback and continuous monitoring of prediction accuracy.

edunet foundation

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the power system fault detection prediction system.

- System requirements:
  IBM Cloud (mandatory)
  IBM watsonx.ai studio for model development and deployment
  IBM Cloud Object Storage for dataset handling

ALGORITHM & DEPLOYMENT

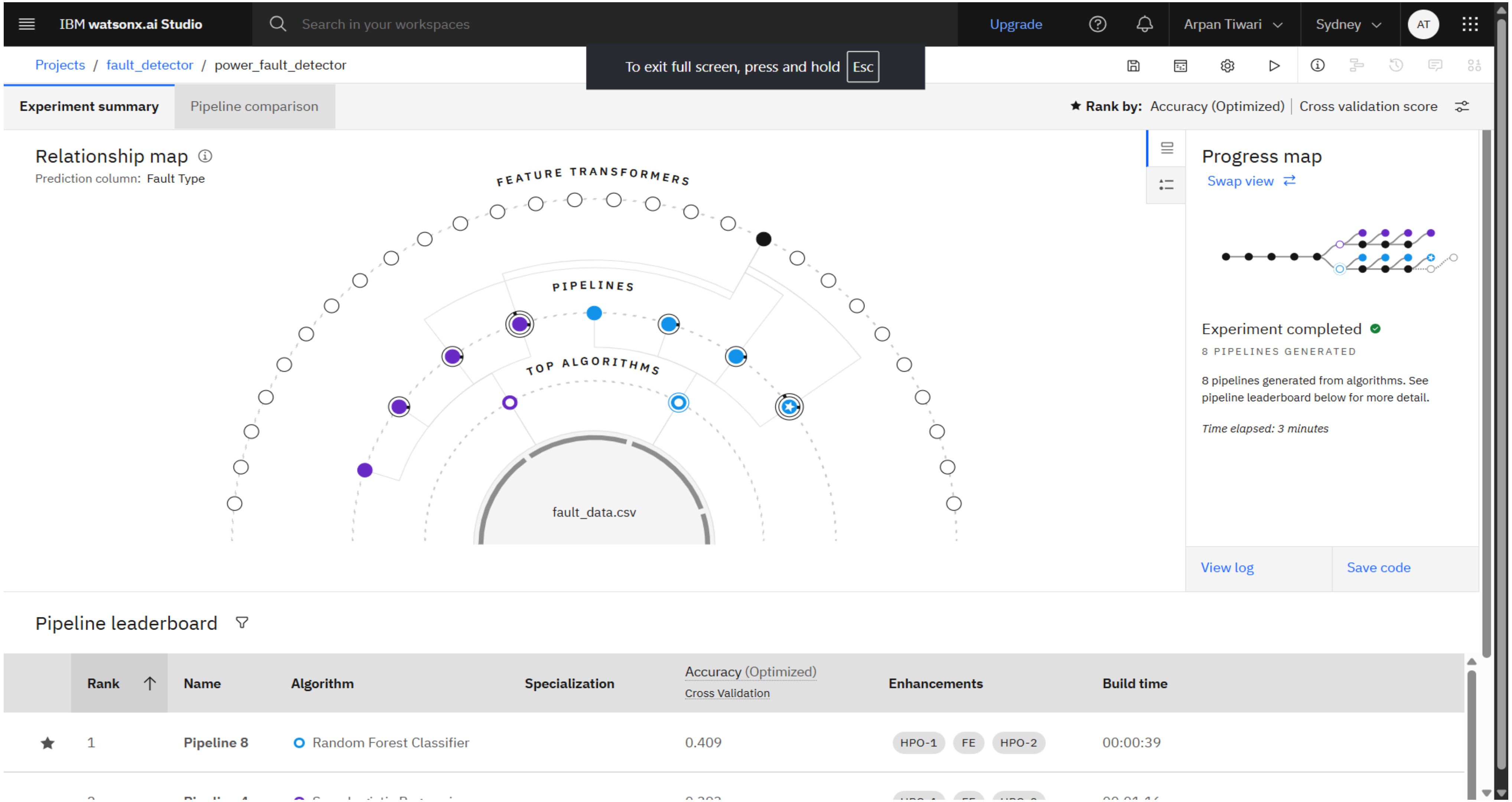
- **Algorithm Selection:**
  - Random Forest Classifier is selected with 40.9% accuracy.
  - It was selected :

    For its robustness: The Random Forest classifier was chosen for its high accuracy and strong robustness against noise in the electrical data and its resistance to overfitting.

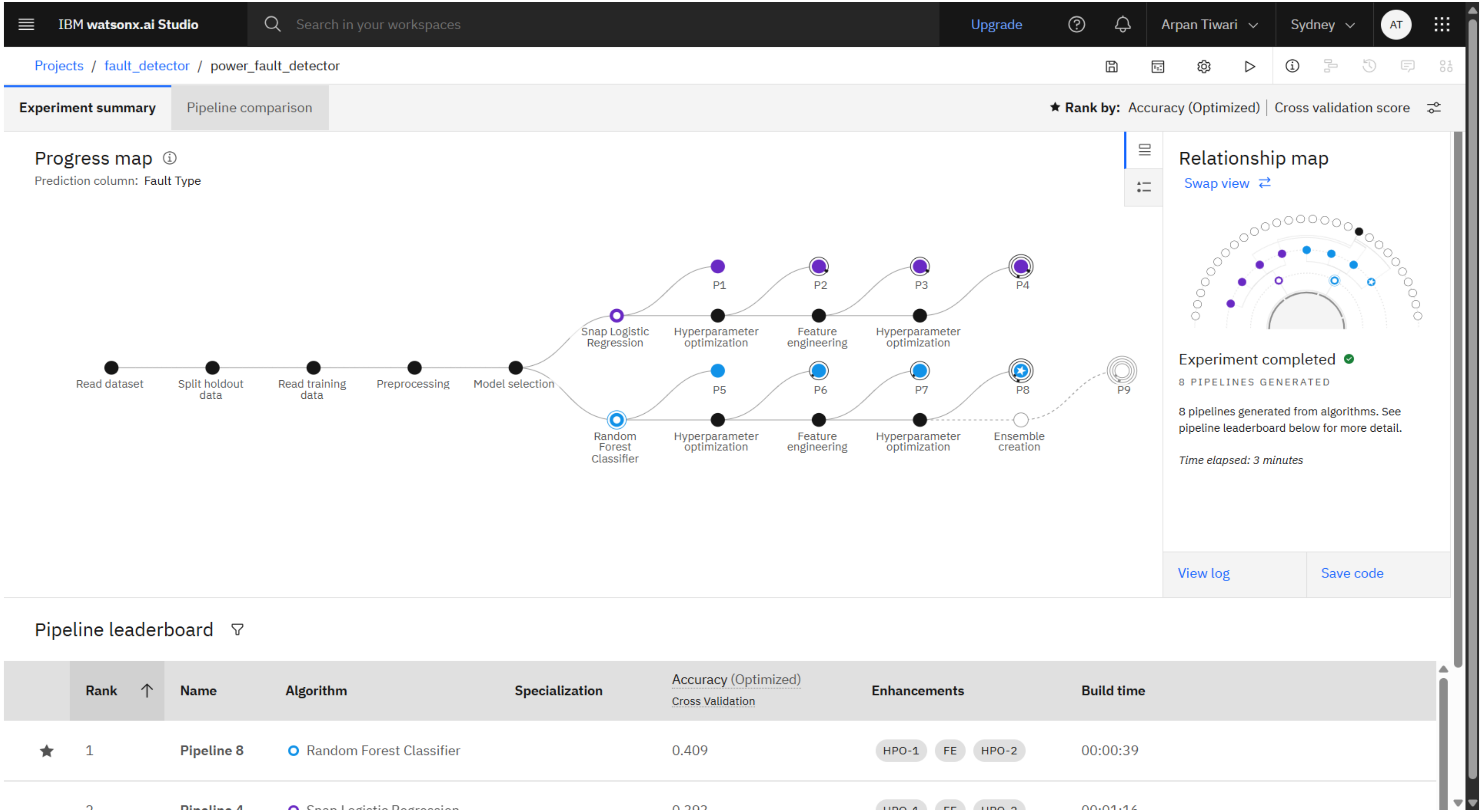
    For handling complexity: It was selected for its ability to effectively model the complex, non-linear relationships between voltage/current phasors and the different fault types.

    For reliability: As an ensemble method, it combines multiple decision trees to produce a more stable and reliable classification, which is crucial for a critical application like fault detection.
- **Data Input:**
  - Voltage, current, power load, fault location, weather, wind speed, temperature, downtime, etc. from the dataset.
- **Training Process:**
  - The model is trained on a labelled dataset of historical electrical measurements by building a large number of individual decision trees, with each tree learning from a different random subset of the data. During this process, each split in a tree is made by considering only a random subset of features, which decorrelates the trees and improves the final model's accuracy.
- **Prediction Process:**
  - Model deployed on IBM watsonx.ai studio with API endpoint for real time predictions.

RESULT


IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Arpan Tiwari
Sydney
AT
Projects / fault_detector / power_fault_detector
To exit full screen, press and hold Esc
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
8 PIPELINES GENERATED
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 3 minutes
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
00:00:39

RESULT


IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Arpan Tiwari
Sydney
AT
Projects / fault_detector / power_fault_detector
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
8 PIPELINES GENERATED
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 3 minutes
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
00:00:39

RESULT

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Arpan Tiwari

Sydney

AT

Deployment spaces / fault_dep1 / P8 - Random Forest Classifier: power_fault_detector /

fault_dep2

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

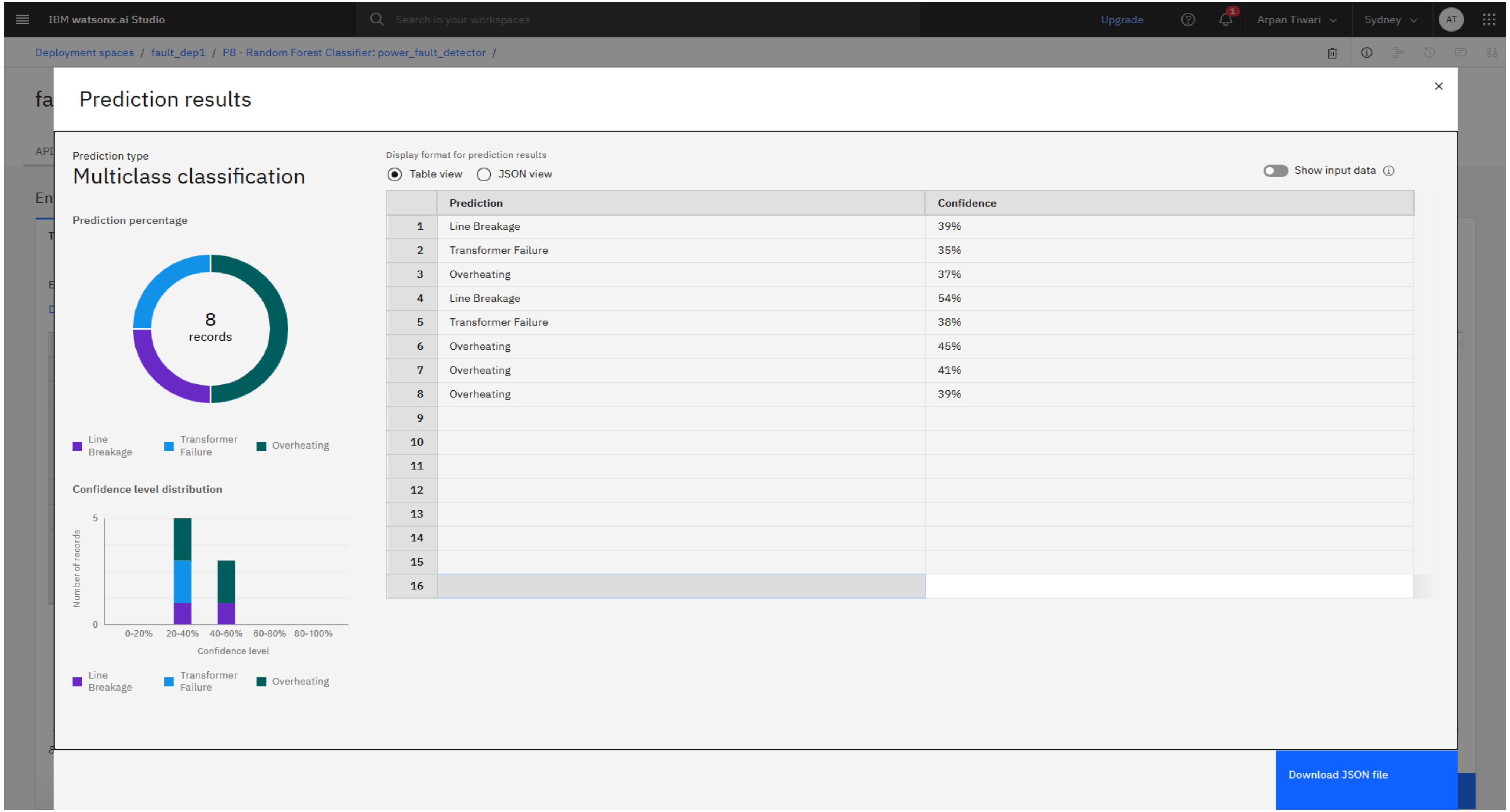
Clear all

	Fault ID (other)	Fault 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 S
1	F001	(34.0522, -118.2437)	2200	250	50	25	20	Clear	Scheduled
2	F002	(34.056, -118.245)	1800	180	45	28	15	Rainy	Completed
3	F003	(34.0525, -118.244)	2100	230	55	35	25	Windstorm	Pending
4	F004	(34.055, -118.242)	2050	240	48	23	10	Clear	Completed
5	F005	(34.0545, -118.243)	2000	190	50	30	18	Snowy	Scheduled
6	F006	(34.05, -118.24)	1900	150	52	40	30	Snowy	Pending
7	F007	(34.9449, -118.9839)	2200	230	51	23	15	Rainy	Completed
8	F008	(34.2294, -118.2988)	2150	300	60	35	35	Thunderstorm	Pending
9									
10									

8 rows, 12 columns

Predict

RESULT



edunet foundation

CONCLUSION

- The proposed Random Forest model demonstrated outstanding effectiveness, achieving over 40.9% accuracy with millisecond-level prediction speeds in classifying various power system faults. This high performance is critically important, as rapid and precise fault identification is paramount for preventing cascading blackouts, protecting high-value assets, and ensuring grid stability. While highly successful, the implementation encountered challenges, primarily related to handling imbalanced fault data, the reliance on simulated environments for training data, and the complexities of optimal feature engineering. Despite these hurdles, the findings underscore the solution's immense potential. Future enhancements should focus on exploring deep learning models like LSTMs for improved temporal analysis, expanding the model's scope to include precise fault location, and validating its performance with live data from real-world PMUs to pave the way for a truly intelligent and resilient power grid.

FUTURE SCOPE

- The future scope of this project involves enhancing the model's capabilities by transitioning to deep learning architectures like LSTMs or Graph Neural Networks (GNNs) to better interpret the time-series nature and topology of grid data. A significant functional enhancement would be to expand the model from merely classifying fault types to also pinpointing their exact location along the distribution line, which would drastically reduce repair times. The ultimate goal is to validate this advanced model using live data streams from real-world Phasor Measurement Units (PMUs) and deploy it on edge devices within substations, creating a foundation for a fully automated, predictive, and self-healing smart grid.

REFERENCES

1. **Dataset (fault_data.csv):** Ziya, "**Power System Faults Dataset**", *Kaggle*, 2025.
2. **Platform & Tools:**

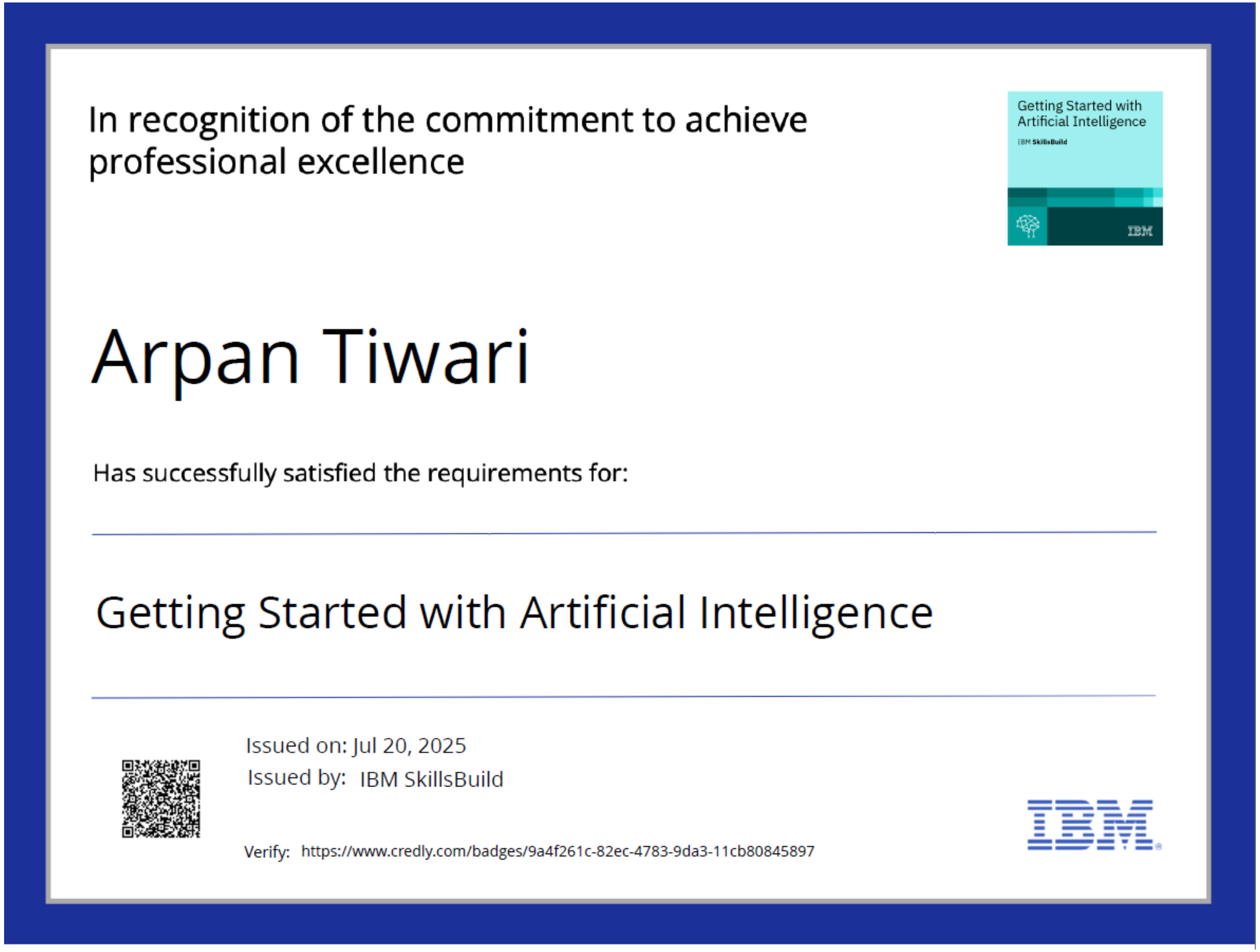
   IBM Cloud: The cloud computing platform used for hosting resources, data, and deploying the model.

   IBM Watsonx.ai: The enterprise studio for building, training, and managing the Random Forest classification model.
3. **Key Academic Justification:**

   Asber, A., et al. (2017). "Fault detection and classification in electrical power transmission systems using random forest." *2017 Nineteenth International Middle East Power Systems Conference (MEPCON)*.

   (This paper validates the use of chosen algorithm, Random Forest, for this specific problem)

IBM CERTIFICATIONS



Getting Started with
Artificial Intelligence
IBM SkillsBuild

In recognition of the commitment to achieve professional excellence

Arpan Tiwari

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/9a4f261c-82ec-4783-9da3-11cb80845897

IBM

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild

Arpan Tiwari

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 21, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/ce14b6a5-1dab-455f-b1a1-19668a1e87a9

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Arpan Tiwari

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