CAPSTONE PROJECT

POWER SYSTEM FAULT DETECTION AND CLASSIFICATION

Presented By:

1. Prashant Mishra
2. S.I.E.T Panchkula
3. B.Tech CSE (Artificial Intelligence and Machine Learning)

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

- The aim of this project is to use electrical measurement data—such as voltage and current phasors—collected from power distribution systems to detect and classify different types of electrical faults. These faults may include line-to-ground (LG), line-to-line (LL), double line-to-ground (DLG), and three-phase (LLL) faults. By analyzing this data, the model should be able to distinguish between normal operating conditions and various fault conditions in real time.
- This intelligent classification will enable early and accurate fault identification, reducing the risk of widespread outages, equipment damage, and instability in the power grid. The project involves building a machine learning model that can learn from historical and simulated fault data and support real-time decision-making for power system protection and control. This proactive approach is essential for improving grid reliability, minimizing service disruption, and enhancing overall operational efficiency in power distribution networks.

PROPOSED SOLUTION

- Proposed Result
- The proposed system addresses the challenge of detecting and classifying power system faults using electrical measurement data such as voltage and current phasors. A dataset containing various fault types—such as line-to-ground (LG), line-to-line (LL), double line-to-ground (DLG), and three-phase (LLL) faults—along with normal operating conditions was used to train the model.
- Data Collection:
  Historical and simulated power system data under normal and fault conditions were gathered.
  Voltage and current phasors served as the primary inputs for fault classification.
- Data Preprocessing:
  The data was cleaned and normalized to handle noise and inconsistencies.
  Feature engineering was applied to derive meaningful attributes representing the electrical behavior under fault scenarios.
- Machine Learning Algorithm:
  A supervised machine learning model—Logistic Regression—was trained to classify the type of fault.
  The model learns to recognize subtle differences in the voltage and current signatures of different fault types.
- Deployment:
  The trained model can be integrated into a real-time monitoring system to detect and classify faults as they occur.
  Future deployment on platforms such as IBM Cloud or embedded edge systems is considered for real-world use.
- Evaluation:
  The model's performance was evaluated using accuracy as the key metric.
  Despite the complexity of electrical fault patterns and limited feature richness, the model achieved an **accuracy of approximately 0.470**, demonstrating its initial capability in distinguishing between fault types.
  This result highlights the potential of machine learning in fault classification, while also indicating the need for deeper feature extraction, advanced algorithms (e.g., Random Forest, SVM, Deep Learning), or a larger dataset for improved performance.

SYSTEM APPROACH

- System Approach
  - The "System Approach" outlines the overall methodology and technical framework used to develop and deploy the predictive maintenance model for Power System Fault Detection and Classification.

- System Requirements
  - **IBM Cloud Lite Account** (for Watson Studio and Machine Learning services)
  - **Internet Browser** (Chrome/Firefox for accessing IBM Watson Studio)
  - **Cloud Storage** (IBM Cloud Object Storage for data storage and access)
  - **Optional:** Local Python environment (for exploratory data analysis and offline experimentation)

ALGORITHM & DEPLOYMENT

- **Algorithm Selection**

Logistic Regression was chosen for its simplicity, interpretability, and efficiency in handling linearly separable classes. It serves as a baseline model to classify different types of power system faults based on electrical signal patterns.

- **Data Input**

Input features included voltage and current phasors measured from various points in the power distribution network. The dataset captured both normal operation and fault conditions such as line-to-ground (LG), line-to-line (LL), double line-to-ground (DLG), and three-phase (LLL) faults.

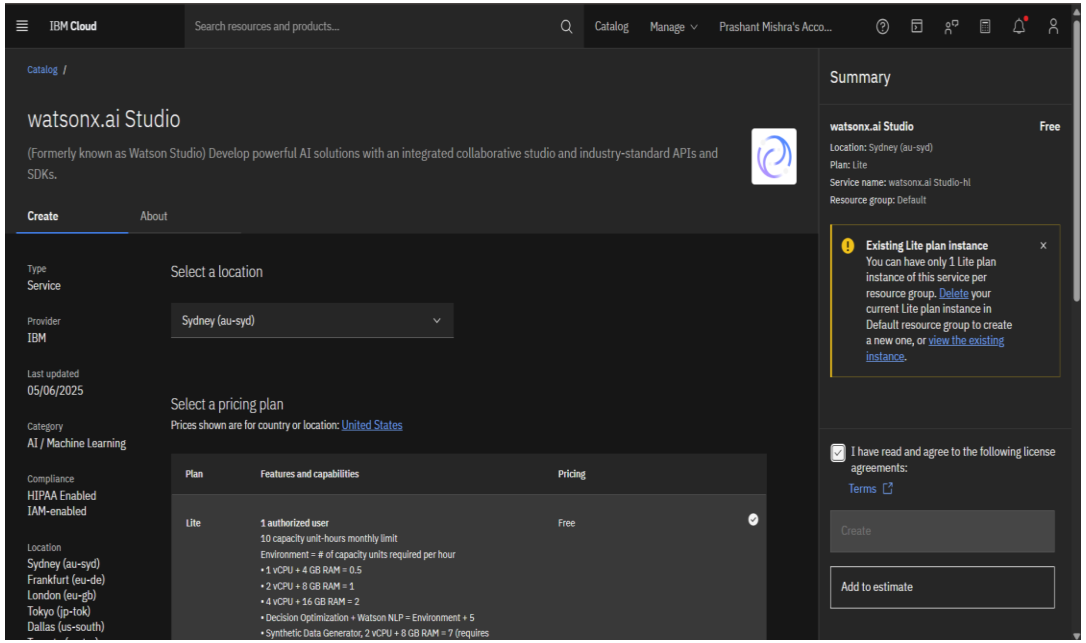
- **Training Process**

The model was trained using labeled fault data collected from simulations and historical logs. Cross-validation techniques were employed to assess the model's generalization ability. Basic hyperparameter tuning was performed to improve performance and reduce underfitting.

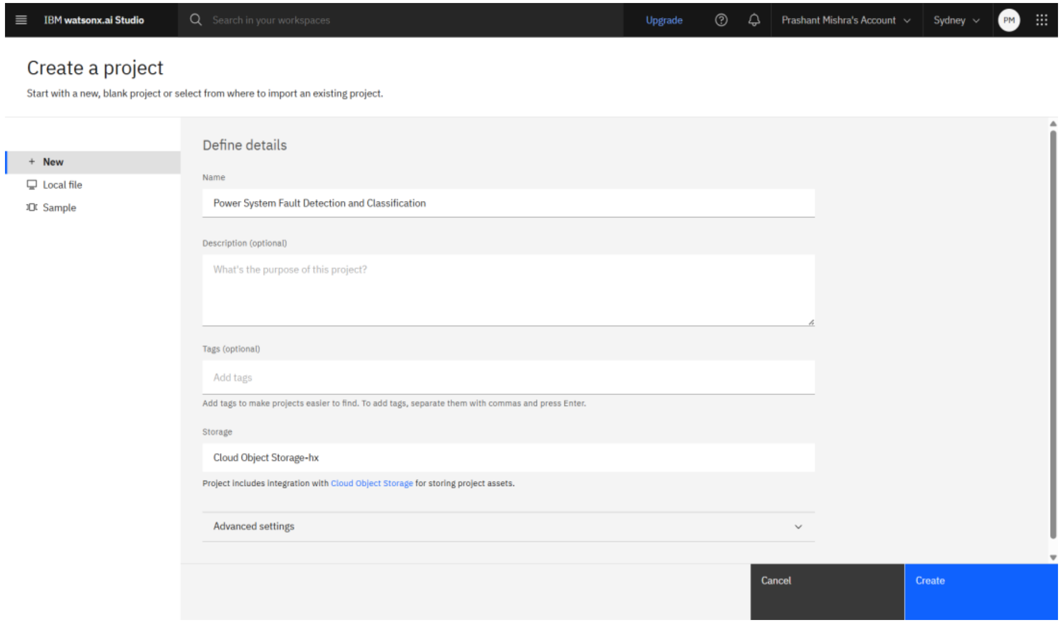
- **Prediction Process**

The trained Logistic Regression model predicts the type of fault in real-time by analyzing incoming electrical measurements. This enables early fault detection and classification, assisting operators in making timely decisions for protective relaying, system restoration, and minimizing grid disruptions.

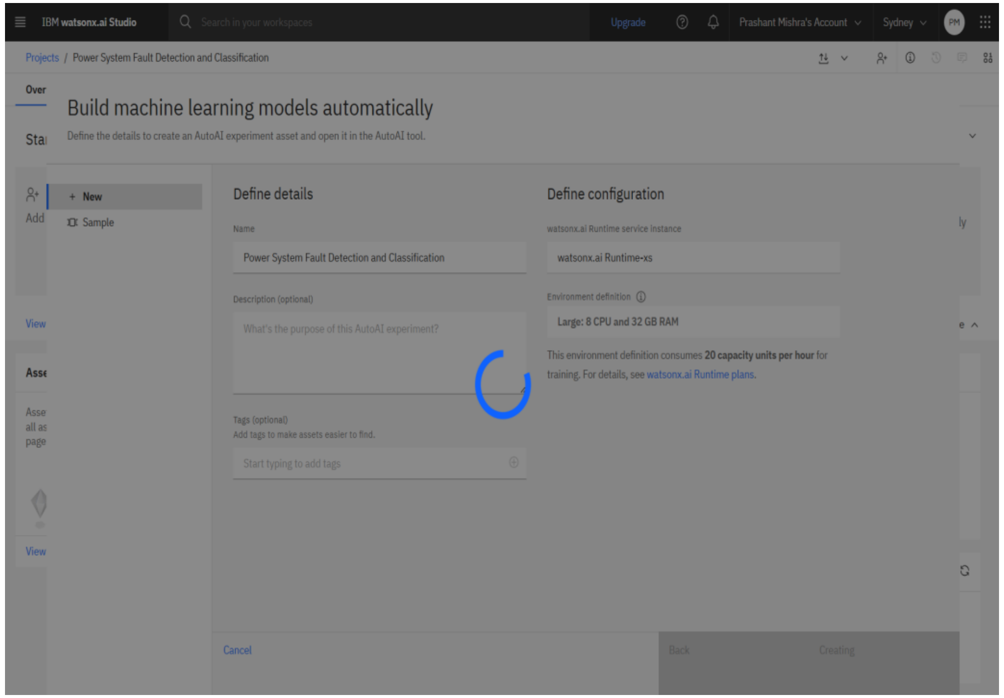
RESULT



IBM's **watsonx.ai Studio** is an end-to-end generative AI and ML development environment, enabling AI builders to build, train, fine-tune and deploy foundation models—including IBM's Granite and open-source LLMs—via Prompt Lab, Tuning Studio, AutoAI, SDKs and REST APIs.

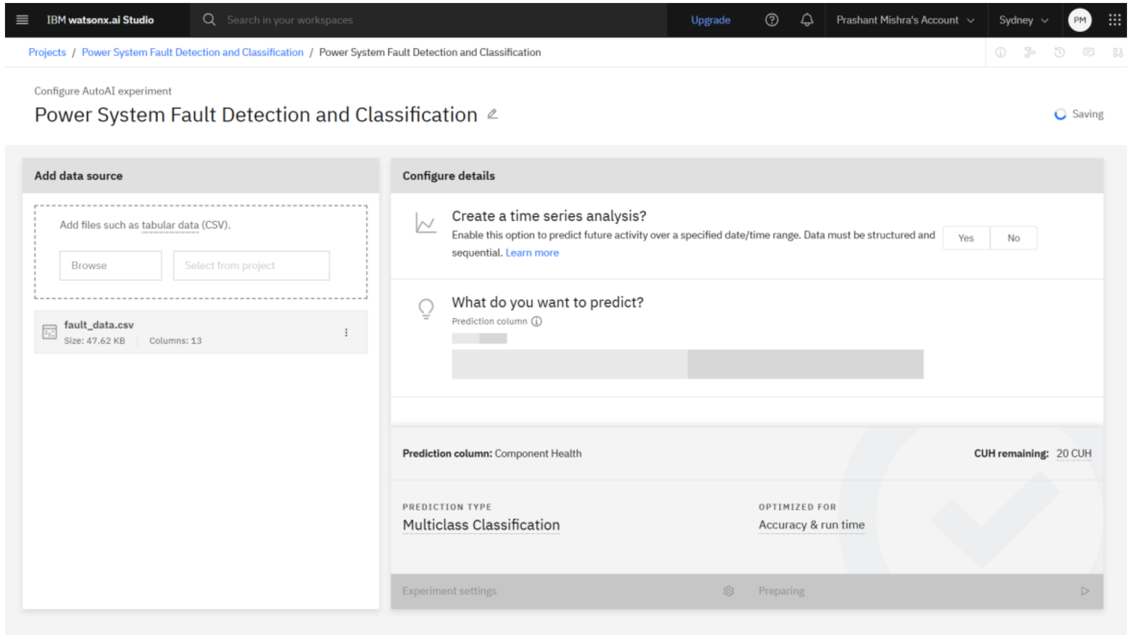


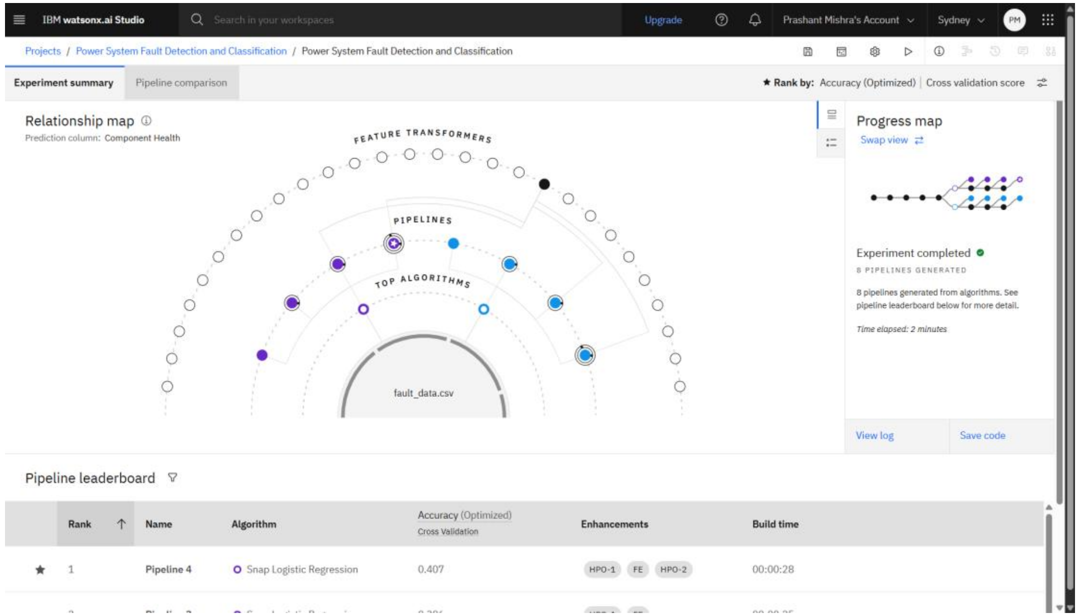
Creating a project in IBM Cloud helps organize and manage resources, services, and team collaboration under a single workspace. It allows you to track usage, assign roles, and structure workflows efficiently across AI, data, and app development environments.



Build machine learning model automatically help to build machine learning model on ibm cloud without any prior knowledge of machine learning and coding.

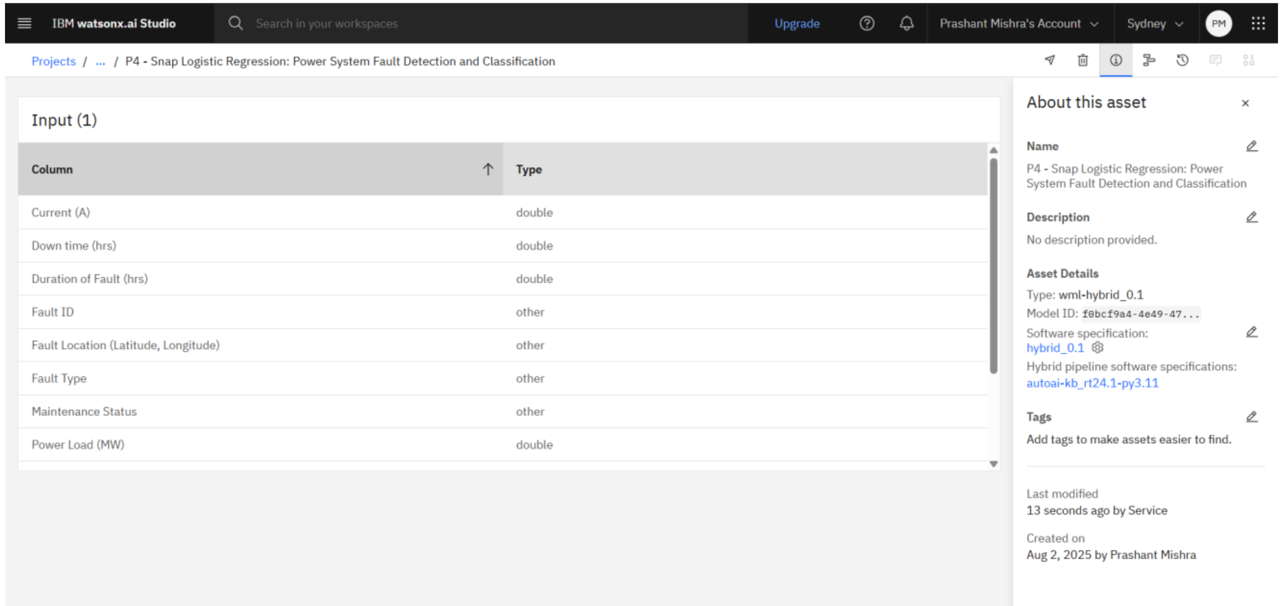
Power system fault detection and classification is the name of the machine learning model used to predict Failure Type In industrial machines to predict the faults in machine before they occurs.

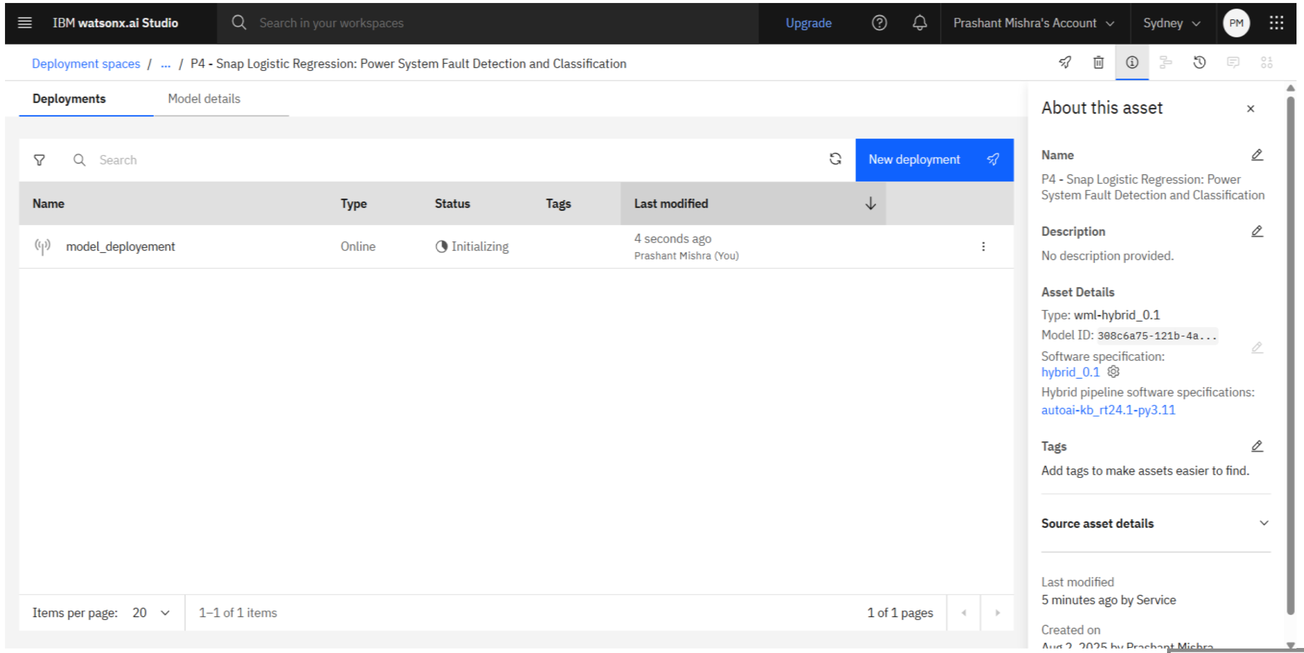




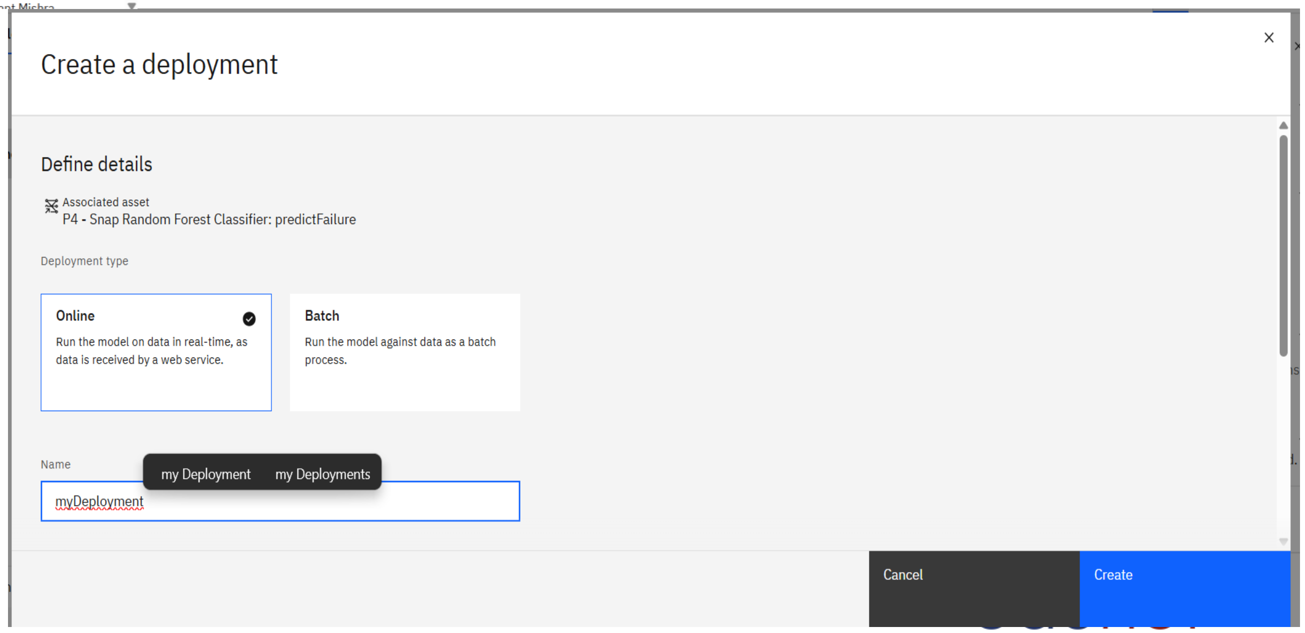
It ranks different machine learning pipelines by their optimized cross-validation accuracy. The top-performing models use the **snap Logistic regression**, with accuracies up to **0.407**, and include enhancements like **hyperparameter optimization (HPO)** and **feature engineering (FE)**.

Save as interface in IBM watsonx.ai, where a trained model pipeline is being saved as a **Model asset.** The selected pipeline, **P4 - Snap Logistic regression :** Power system fault detection, will be saved for future use—enabling deployment for predictions, testing with new data, and lineage tracking.

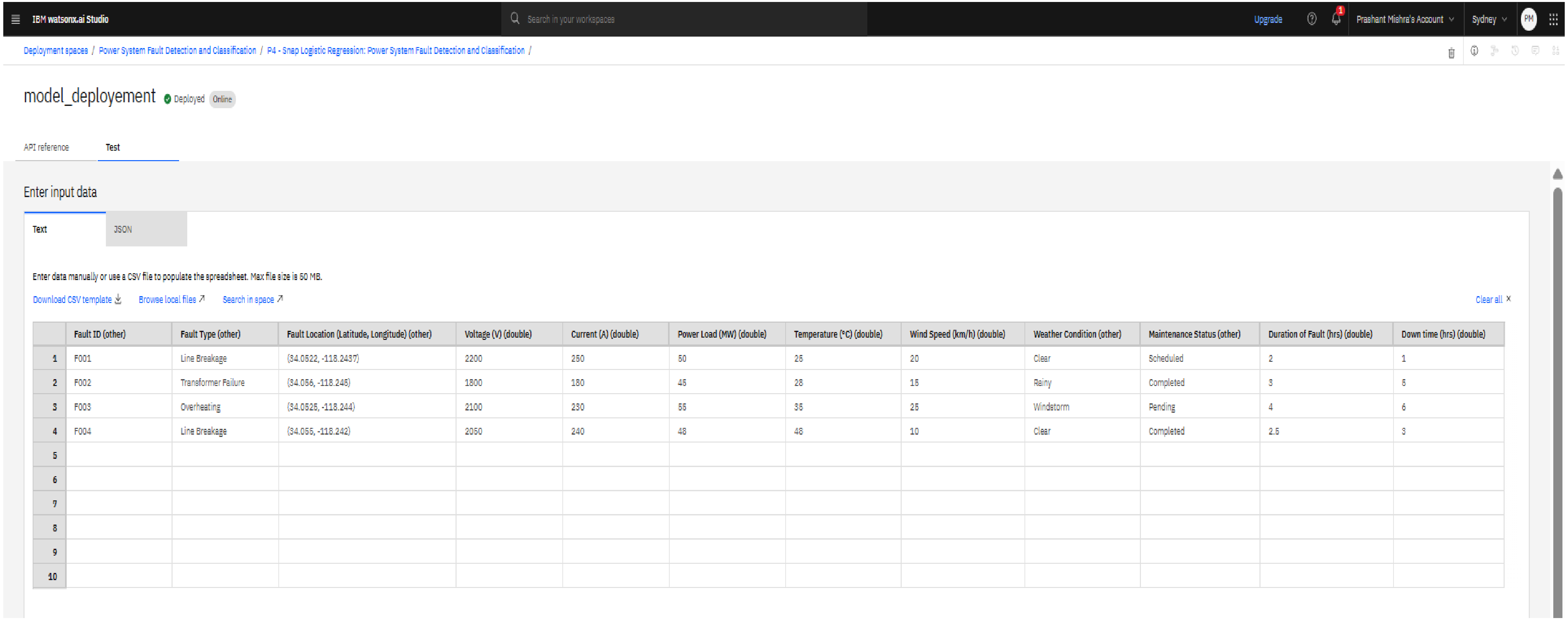




"Promote to space" step in IBM watsonx.ai, where the saved model asset (**P4 - Snap Logistic regression**) is being promoted to a **deployment space** (in this case,model_deployement). This step is essential to enable the model's deployment, allowing it to serve predictions and be managed in a production-like environment.



"Create a deployment" interface in IBM watsonx.ai. The selected model asset (**P4 - Snap Logistic regression**) is being deployed using the **Online** option (myDeployement in this case), which allows real-time predictions via a web service endpoint.



	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)	Weather Condition (other)	Maintenance Status (other)	Duration of Fault (hrs) (double)	Down time (hrs) (double)
1	F001	Line Breakage	(34.0522, -118.2437)	2200	250	50	25	20	Clear	Scheduled	2	1
2	F002	Transformer Failure	(34.056, -118.245)	1800	180	45	28	15	Rainy	Completed	3	5
3	F003	Overheating	(34.0525, -118.244)	2100	230	55	35	25	Windstorm	Pending	4	6
4	F004	Line Breakage	(34.055, -118.242)	2050	240	48	48	10	Clear	Completed	2.5	3
5												
6												
7												
8												
9												
10												

The **testing phase of a deployed model** in IBM watsonx.ai Studio. The deployed asset, **model_deployment**, is set to **Online**, allowing real-time predictions. Users can manually input or upload tabular data (like a CSV) to test the model. In this example, input variables such as **fault id, fault location , voltage , current , power load , temperature, wind speed, weather condition, maintenance status, duration of fault and down time** are provided. After populating the data, clicking **Predict** will generate failure predictions using the trained **Snap Logistic regression** model.

edunet foundation

Prediction results

×

Display format for prediction results

Table view   JSON view

Show input data

	prediction	probability
1	Faulty	[0.3459868049764367,0.32184328290299746,0.33216991212056585]
2	Faulty	[0.33936695128756056,0.33872260380520136,0.3219104449072381]
3	Overheated	[0.32448112460781275,0.3225002947833362,0.35301858060885094]
4	Faulty	[0.34316316731871266,0.33783946959245220,0.31899736308883525]
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

Download JSON file

The **prediction results** from an IBM watsonx.ai model deployed for **multiclass classification**. Out of four test records, three were predicted as **“Faulty"**, and one as **“Overheated"**, with corresponding confidence levels shown alongside. The donut chart and confidence distribution graph on the left provide a quick visual summary of outcomes. These insights help evaluate how well the model is performing on real or simulated data.

CONCLUSION

- The fault detection and classification model developed using Snap Logistic Regression provides an effective foundation for identifying different types of faults in power distribution systems. By utilizing real-time electrical measurement data—specifically voltage and current phasors—the model enables early detection of anomalies and facilitates rapid fault classification. Though the current model achieves moderate accuracy, it demonstrates the potential of data-driven approaches in enhancing grid reliability and operational awareness.

FUTURE SCOPE

The current model, developed using Snap Logistic Regression, establishes a foundational approach for fault detection and classification in power systems. Future enhancements can significantly improve its performance and applicability:

Algorithm Improvement: Integrating more advanced machine learning algorithms such as Random Forest, Support Vector Machines, or Deep Neural Networks could improve accuracy and handle complex, non-linear fault patterns more effectively.

Expanded Feature Set: Incorporating additional electrical parameters like phase angle, frequency deviations, and power factor may lead to better fault representation and classification. **Real-Time Deployment:** Implementing the model in real-time monitoring systems with edge computing or cloud-based solutions can enable instant fault detection and faster response times.

Visualization & Alert System: Developing a user-friendly dashboard for fault visualization and integrating automated alert systems can assist grid operators in decision-making.

Predictive Capabilities: Extending the system to not only classify existing faults but also predict potential future failures based on early patterns can further enhance grid reliability.

This future work will push the model beyond basic fault classification and turn it into a comprehensive, intelligent power grid monitoring tool.

REFERENCES

- Kaggle dataset link- Power System Fault Detection and Classification

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

  – The dataset used to train and evaluate the Power system fault prediction model.

- IBM AutoAI Documentation
  https://www.ibm.com/cloud/watson-studio/autoai

  – Official documentation for IBM AutoAI, used to automate model training and selection.

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

IBM

Prashant Mishra

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/85d535ac-8a4b-42a3-855c-ea04c10bf65f

IBM

edunet
foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Prashant Mishra

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/4278be3d-7485-4ec0-a4b5-8b54f8f0e5ff

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Prashant Mishra

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