CAPSTONE PROJECT

POWER SYSTEMFAULT DETECTION AND CLASSIFICATION USING MACHINE LEARNING

Presented By:
Mohd Shahnawaz Ashraf
Noida Institute of Engineering and Technology – CSE - AIML

OUTLINE

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

PROBLEM STATEMENT

Faults in power systems like Line-to-Ground (L-G), Line-to-Line (L-L), and 3-phase faults can damage equipment and cause blackouts.

There is a need for a system that can **automatically detect and classify these faults** using voltage and current phasor data.

The goal is to build a machine learning model to accurately identify fault types for maintaining grid stability.

PROPOSED SOLUTION

- The proposed system aims to address the challenge of detecting and classifying faults in a power distribution system using machine learning. The solution leverages voltage and current data to predict whether the system is operating normally or experiencing a specific type of fault. The system is designed to enable fast, accurate fault identification, enhancing the stability and reliability of the power grid.
- Data Collection:
  - Use the publicly available dataset from Kaggle which includes labeled electrical measurements under different fault and normal operating conditions.
  - Focused columns include: Air temperature, Process temperature, Rotational speed, Torque, Tool wear, and Type.
- Data Preprocessing:
  - **Removed irrevelant columns from the dataset like Id's and Names.**
  - **Performed data refinement using IBM Watsonx.ai Data Refinery**
- Machine Learning Algorithm:
  - Use IBM Watsonx.ai's AutoAI tool to automatically generate pipelines and selecting the best-performing pipeline (Snap Random Forest Classifier)
  - **Multiple Algorithms evaluated including Decision Trees and Random Forest Classifier.**
- Deployment:
  - Use the Watsonx.ai interface to test the model with live inputs and visualize the prediction results.
  - Deploy the best model on IBM Cloud using the online deployment option. Expose REST API endpoints for real-time testing and predictions.
- Evaluation:
  - Assess model performance based on accuracy, F1-score, precision, and recall. Top model achieved 99.4% accuracy with high precision across all classes.
  - Predictions successfully distinguish between "No Failure" and various fault types with high confidence.

SYSTEM APPROACH

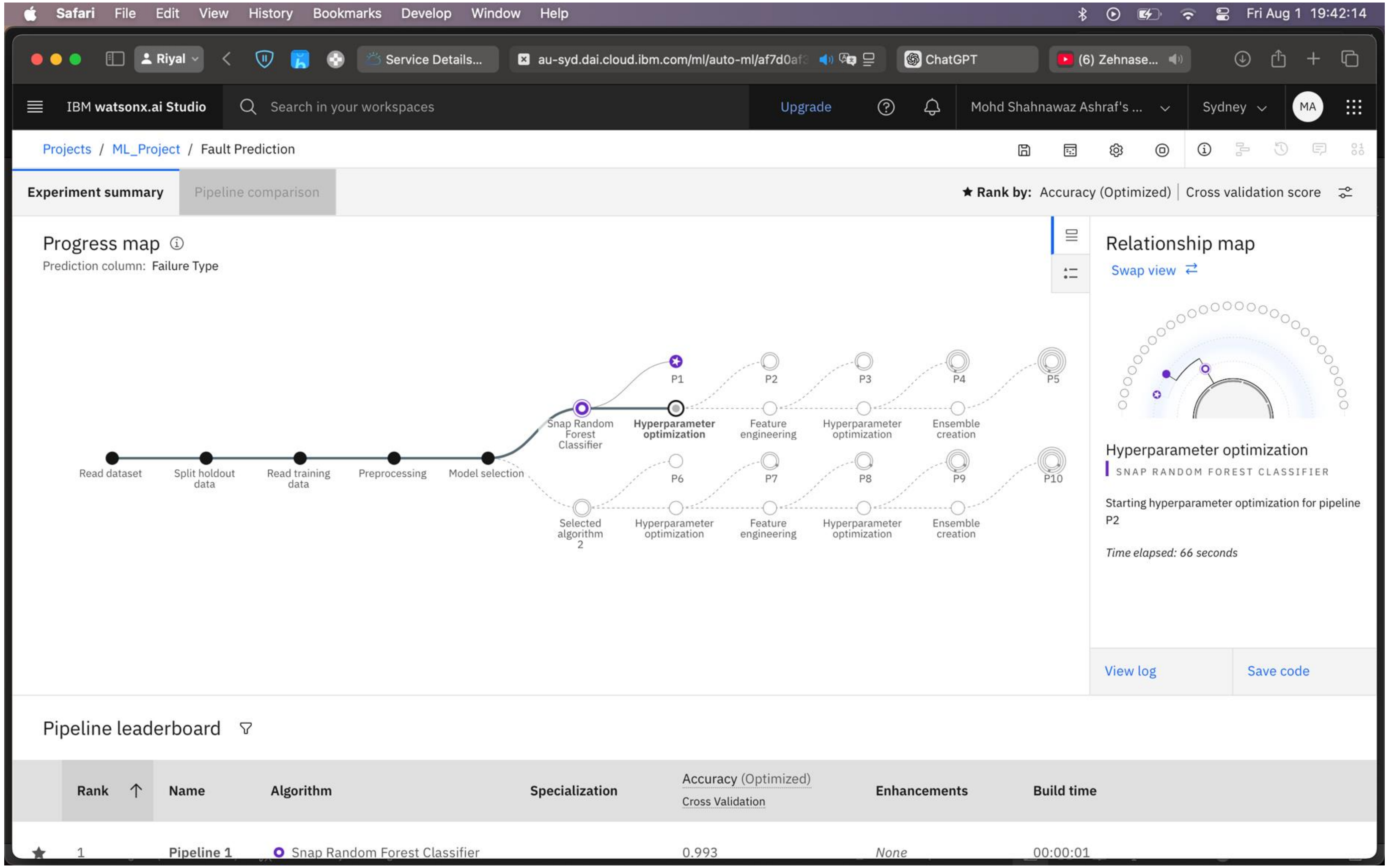
- System requirements
  - A machine learning environment with support for AutoAI workflows
  - Cloud platform for model training, deployment, and testing (IBM Cloud Lite)
  - Tools for data refinement and visualization (Data Refinery)
  - Internet access for dataset retrieval from Kaggle

- Library and Platform Stack:
  - **Platform**: IBM Watsonx.ai Studio
  - **Model Builder:** IBM AutoAI (AutoML tool within Watsonx.ai)
  - **Data Refinement**: IBM Data Refinery (within Watsonx.ai)
  - **Deployment & Testing**: IBM Cloud Deployment Space
  - **Underlying Tech**: Python-based ML backend using scikit-learn inside AutoAI

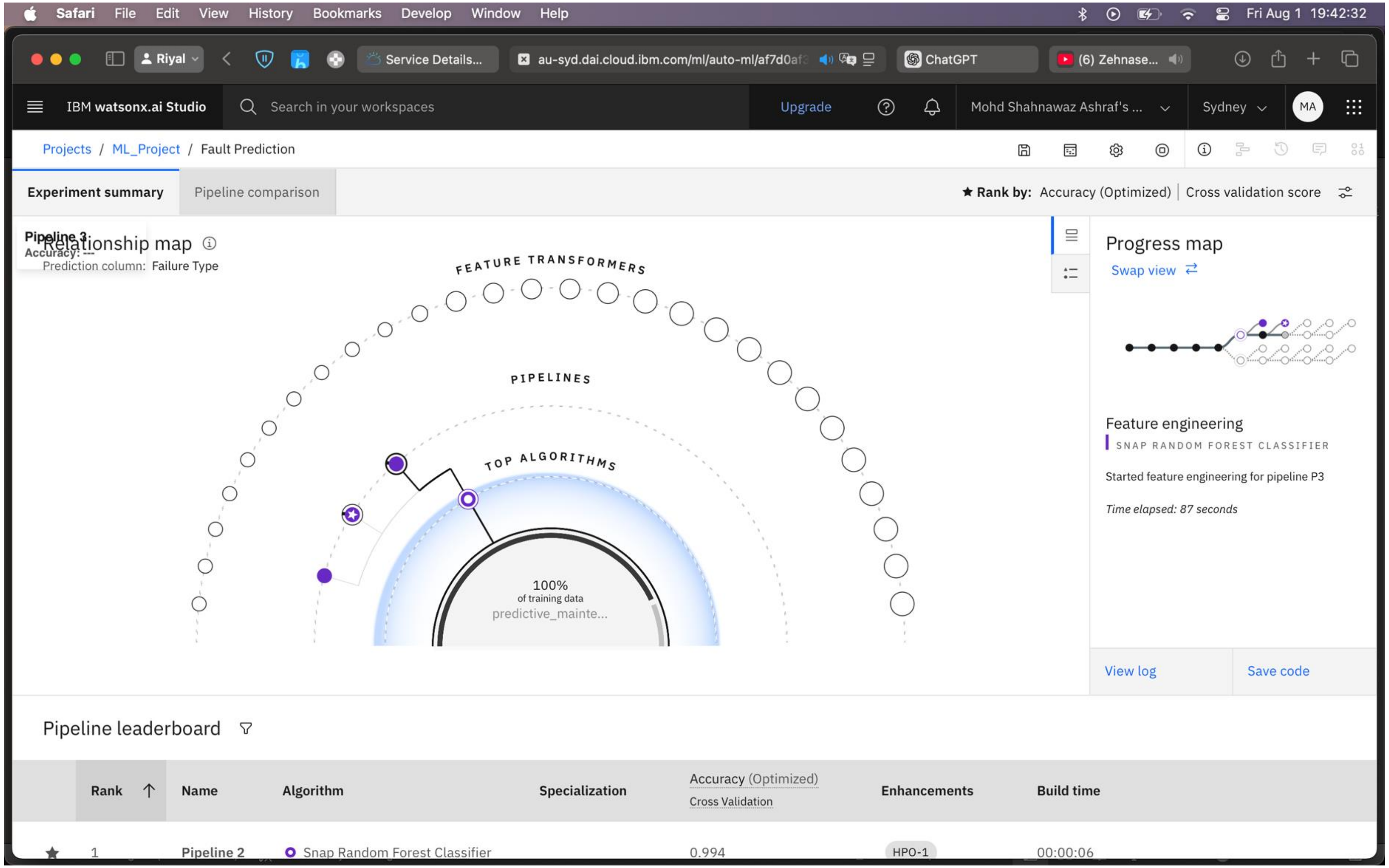
ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  - The selected algorithm is the **Snap Random Forest Classifier**, automatically chosen by IBM Watsonx.ai AutoAI based on its superior accuracy and robustness.
    Random Forest was preferred due to its ability to handle high-dimensional input data and its proven effectiveness in multiclass classification problems like fault detection
- **Data Input:**
  - The algorithm takes the following as the input – **Type** (String/Categorical), **Air Temperature** (Decimal), **Process Temperature** (Decimal), **Rotational Speed** (Integer), **Torque [Nm]** (Decimal), **Tool Wear [min]** (Integer).
- **Training Process:**
  - The dataset was split into training and holdout sets using IBM AutoAI's default split. AutoAI automatically performed:
    - Feature Engineering
    - Hyperparameter Optimization
    - Model Selection using Cross-Validation
      - Each pipeline was built and validated independently, and the best-performing model was Pipeline 2 (Snap Random Forest Classifier) with cross-validated performance metrics.
- **Prediction Process:**
  - In the test phase, the model correctly classified inputs with up to **100% confidence**, demonstrating strong generalization.

ALGORITHM & DEPLOYMENT

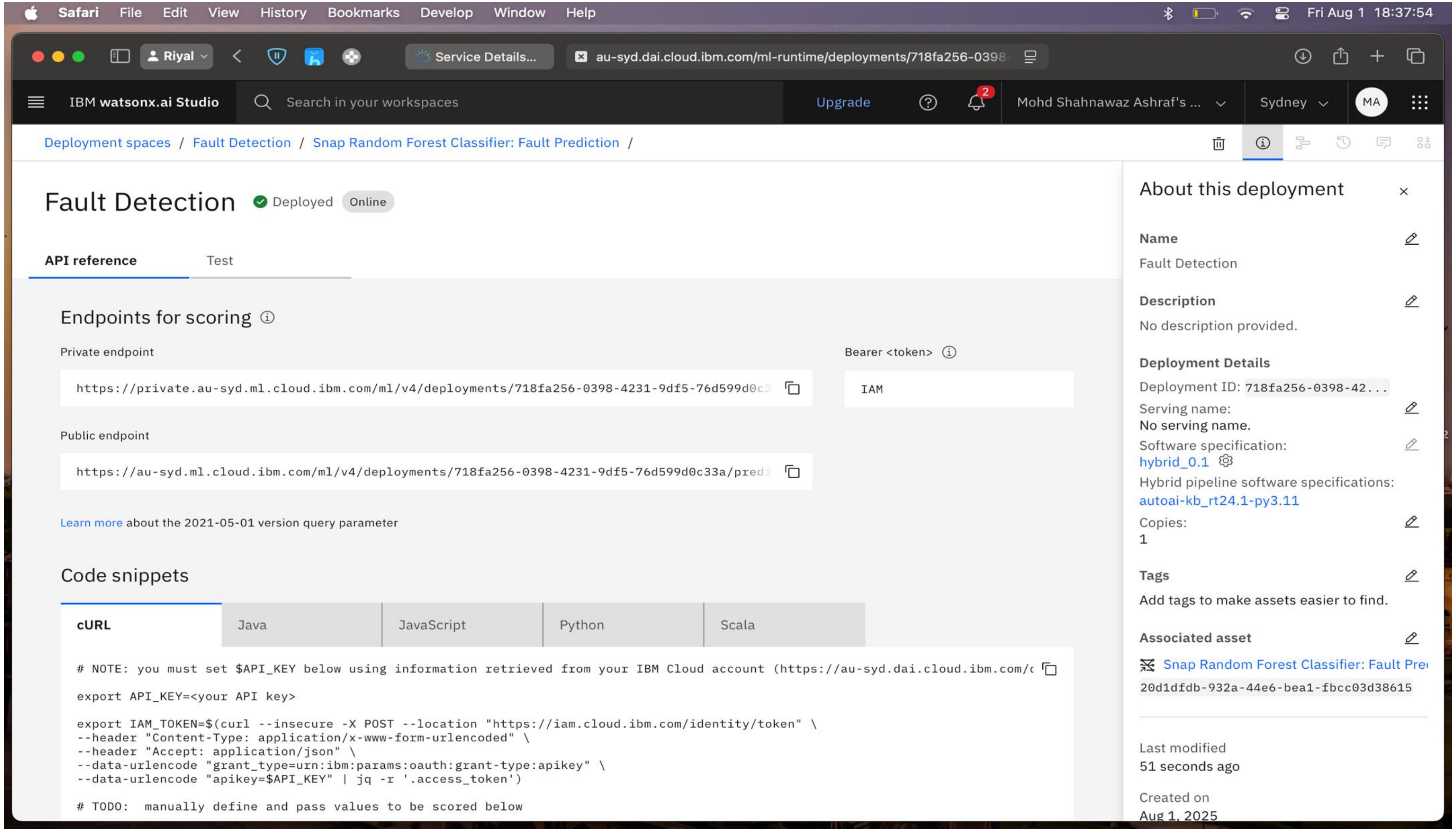


ALGORITHM & DEPLOYMENT



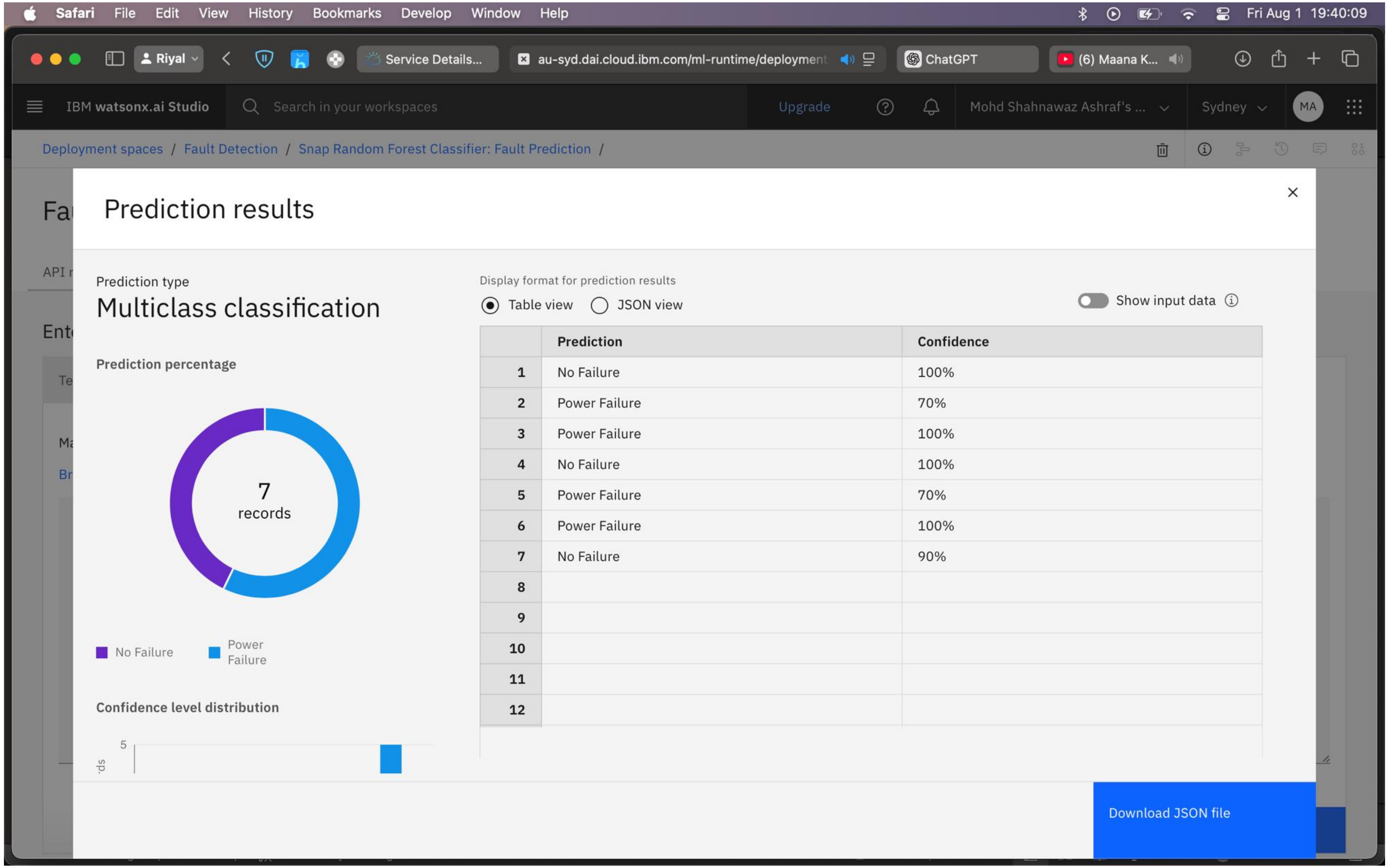
edunet foundation

ALGORITHM & DEPLOYMENT



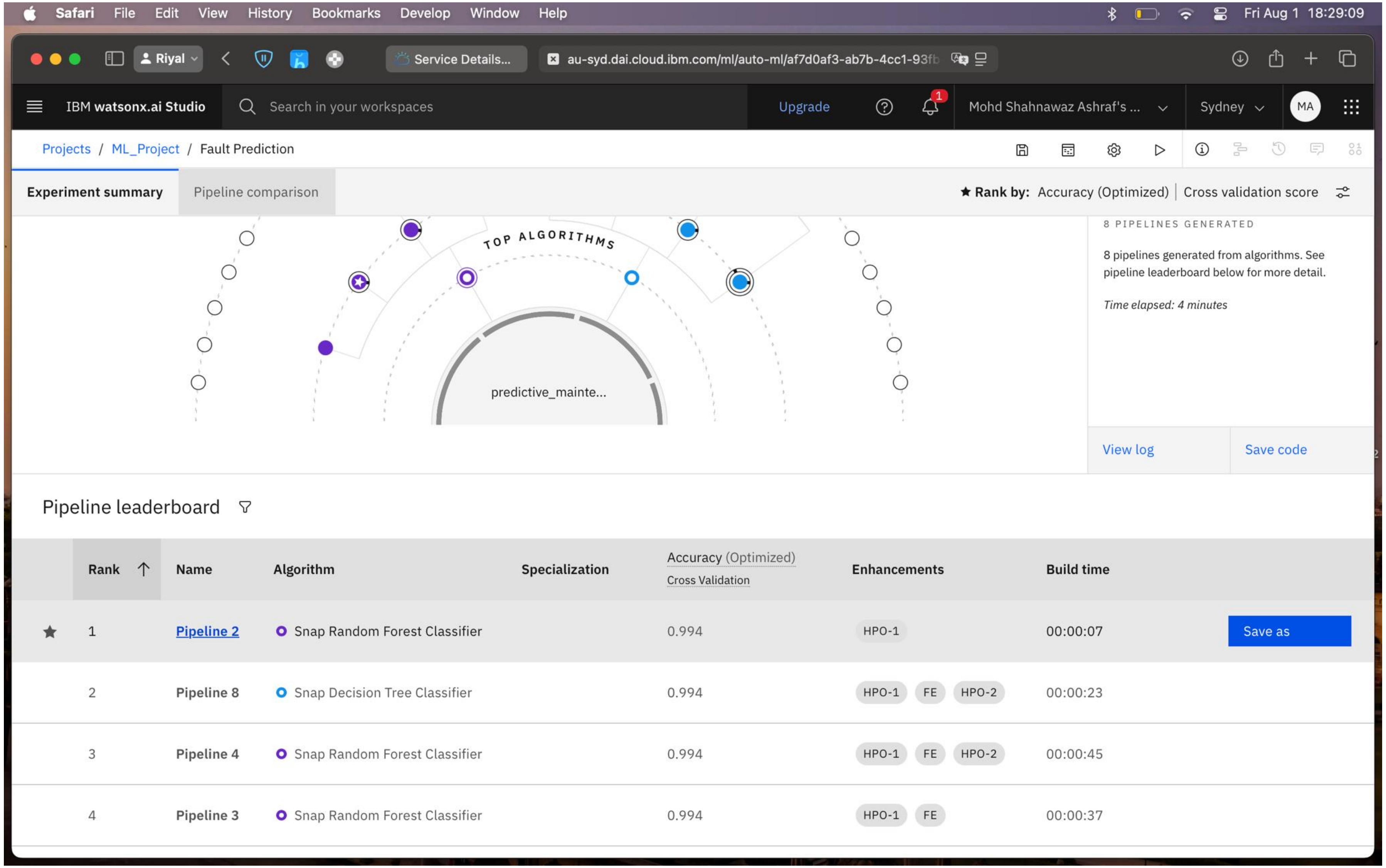
edunet foundation

RESULT

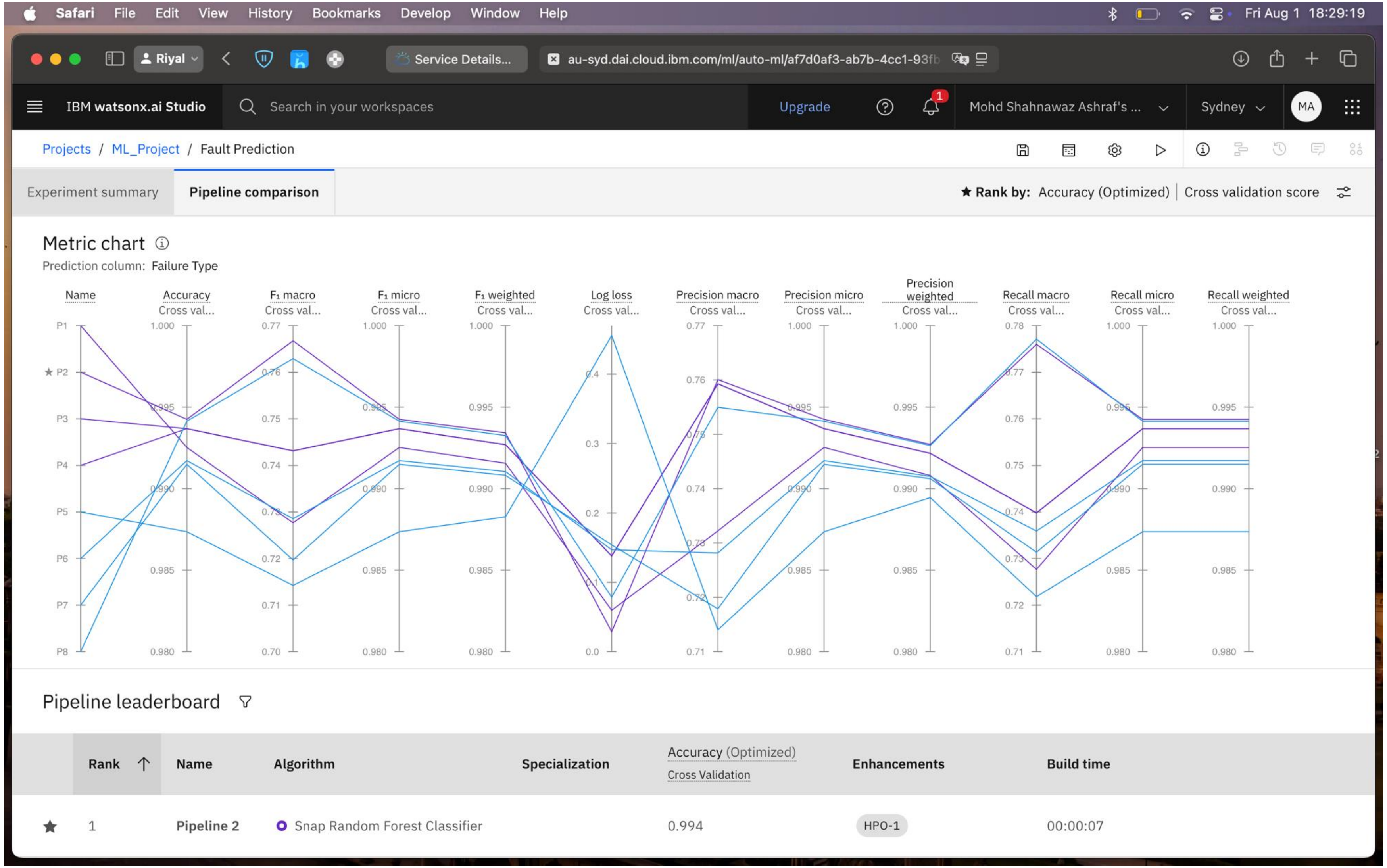


edunet foundation

RESULT


Safari File Edit View History Bookmarks Develop Window Help
Fri Aug 1 18:29:09
Riyal
Service Details...
au-syd.dai.cloud.ibm.com/ml/auto-ml/af7d0af3-ab7b-4cc1-93fb
IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Mohd Shahnawaz Ashraf's ...
Sydney
MA
Projects / ML_Project / Fault Prediction
Experiment summary
Pipeline comparison
★ Rank by: Accuracy (Optimized) | Cross validation score
TOP ALGORITHMS
predictive_mainte...
8 PIPELINES GENERATED
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 4 minutes
View log
Save code
Pipeline leaderboard
Rank | Name | Algorithm | Specialization | Accuracy (Optimized) Cross Validation | Enhancements | Build time
1 | Pipeline 2 | Snap Random Forest Classifier | | 0.994 | HPO-1 | 00:00:07 | Save as
2 | Pipeline 8 | Snap Decision Tree Classifier | | 0.994 | HPO-1 FE HPO-2 | 00:00:23
3 | Pipeline 4 | Snap Random Forest Classifier | | 0.994 | HPO-1 FE HPO-2 | 00:00:45
4 | Pipeline 3 | Snap Random Forest Classifier | | 0.994 | HPO-1 FE | 00:00:37

RESULT



Safari File Edit View History Bookmarks Develop Window Help
Fri Aug 1 18:29:19
Riyal
Service Details...
au-syd.dai.cloud.ibm.com/ml/auto-ml/af7d0af3-ab7b-4cc1-93fb
IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Mohd Shahnawaz Ashraf's ...
Sydney
MA
Projects / ML_Project / Fault Prediction
Experiment summary
Pipeline comparison
Rank by: Accuracy (Optimized) | Cross validation score
Metric chart
Prediction column: Failure Type
Name
Accuracy Cross val...
F1 macro Cross val...
F1 micro Cross val...
F1 weighted Cross val...
Log loss Cross val...
Precision macro Cross val...
Precision micro Cross val...
Precision weighted Cross val...
Recall macro Cross val...
Recall micro Cross val...
Recall weighted Cross val...
P1
P2
P3
P4
P5
P6
P7
P8
Pipeline leaderboard
Rank
Name
Algorithm
Specialization
Accuracy (Optimized)
Cross Validation
Enhancements
Build time
1
Pipeline 2
Snap Random Forest Classifier
0.994
HPO-1
00:00:07

CONCLUSION

- The project successfully demonstrated the use of machine learning, specifically IBM Watsonx.ai's AutoAI, for accurate and efficient classification of power system faults. By training the model on a labeled dataset of electrical parameters such as voltage, current, torque, and temperature, the system was able to distinguish between normal conditions and fault scenarios with a high degree of precision.
- The proposed solution proved effective, with the best-performing model (Snap Random Forest Classifier) achieving **99.4% accuracy**. This high accuracy enables near real-time identification of critical faults like Line-to-Ground or Power Failures, which is essential for maintaining power grid reliability and minimizing downtime.
- Accurate fault detection is crucial in electrical systems, just as accurate bike count predictions are vital for managing urban transportation. This system demonstrates how AI can play a key role in creating smarter, more resilient power infrastructure.

FUTURE SCOPE

- The system can be enhanced and expanded in several ways:
  - **Real-Time Data Integration**: Incorporate IoT sensor data from the grid for real-time fault detection.
  - **Algorithm Improvements**: Explore deep learning models like LSTM or ensemble methods for better accuracy.
  - **Wider Deployment**: Extend the model to support multiple regions or substations with different operating conditions.
  - **Edge Computing**: Deploy the model on edge devices for faster local predictions without relying on cloud latency.
  - **Scalability**: Integrate with dashboard tools for centralized fault monitoring across multiple sites.

REFERENCES

- **Ziya Uddin (2020).**
  Power System Faults Dataset – Kaggle.
  https://www.kaggle.com/datasets/ziya07/power-system-faults-dataset
- **IBM Watsonx.ai Documentation.**
  Build, train, and deploy machine learning models using AutoAI.
  https://www.ibm.com/docs/en/watsonx

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Mohd Shahnawaz Ashraf

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/51490873-faa5-4b64-91a0-6e095ecd165e

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Mohd Shahnawaz Ashraf

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/a0cd73b1-7a85-4325-9045-c18a735ea2a3

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Mohd Shahnawaz Ashraf

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