CAPSTONE PROJECT

PREDICTIVE MAINTENANCE OF INDUSTRIAL MACHINERY

Presented By:
Parnika Mathur
Institute of Information Technology and Management, Janakpuri
BCA (Data Science)

edunet
foundation

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

Problem Statement No. - 39

Develop a predictive maintenance model for a fleet of industrial machines to anticipate failures before they occur. This project will involve analyzing sensor data from machinery to identify patterns that precede a failure. The goal is to create a classification model that can predict the type of failure (e.g., tool wear, heat dissipation, power failure) based on real-time operational data. This will enable proactive maintenance, reducing downtime and operational costs.

PROPOSED SOLUTION

The proposed solution is a machine learning-based system that predicts different types of equipment failures (e.g., tool wear, heat dissipation issues, power failure) before they occur, using real-time sensor data from industrial machines. This helps enable proactive maintenance and minimize unplanned downtime.

- **Data Collection & Input :** Historical sensor data from machines was collected into a Kaggle dataset, including parameters like air temperature, process temperature, rotational speed, torque, and tool wear.
- **Data Preprocessing :** Cleaned data, handled missing values, and applied basic feature engineering to enhance model input quality.
- **Model Development :** IBM Cloud AutoAI was used to build, test, and rank multiple classification models. The platform handled feature transformations, model selection (e.g., Decision Tree Random Forest), and hyperparameter optimization.
- **Deployment :** Deployed the best model via IBM Cloud, making real-time predictions accessible through an API endpoint.
- **Evaluation :** Assessed model performance using metrics like accuracy, precision, recall, and F1-score for multiclass classification.
- **Outcome :** The system helps identify potential failures in advance, supports timely maintenance actions, and improves machine reliability.

edunet
foundation

SYSTEM APPROACH

System Requirements

- **Cloud Platform:** IBM Cloud
- **Service Used:** IBM Watsonx.ai Studio, IBM Watsonx.ai Runtime with AutoAI and Cloud Storage Object of free specs minimum
- **Input Data:** Real-time sensor data from industrial machines (sourced from Kaggle)
- **Model Type:** Multiclass Classification Model (to predict type of failure)
- **Deployment:** IBM Cloud Deployment Space

Libraries used internally by IBM Cloud

- scikit-learn – For preprocessing, model training, and evaluation
- pandas – For data handling
- numpy – For numerical operations
- matplotlib / seaborn – For visualizations during AutoAI model analysis and analysing the predictions.

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:** The AutoAI platform on IBM Cloud evaluates and selects the best machine learning algorithm for multiclass classification tasks. In this case, algorithms like **Decision Tree, Random Forest, Gradient Boosting** were auto-tested. These are well-suited for handling structured sensor data and classifying failure types. The final selection is based on accuracy, F1-score, and AutoAI's pipeline optimization process.
- **Data Input:** The input features include sensor readings such as Air Temperature, Process Temperature, Rotational Speed, Torque, Tool Wear and the **Target is Failure type**
- **Training Process:** The historical sensor data is split into training and testing sets. AutoAI automatically performs Data preprocessing and scaling, Feature selection, Model training and evaluation, Cross-validation to prevent overfitting and Hyperparameter tuning for optimal performance
- **Prediction Process:** Once deployed, the trained model receives real-time sensor data as input. It processes the incoming data through the optimized pipeline and predicts the specific failure class (e.g., power failure). The predictions allow seamless integration with industrial monitoring systems for proactive maintenance alerts.

ALGORITHM & DEPLOYMENT

- A **decision tree** actually looks like a tree, where each internal **node** represents a test on a feature, each **branch** shows the outcome of that test, and each **leaf** node gives the final result or decision. They are used for classification and regression tasks because they are easy to understand and interpret. They work by splitting data into smaller groups based on the most important features, helping to reveal patterns and make accurate predictions.

- A **Random Forest** is an advanced machine learning method that builds on decision trees. It creates a "forest" of many decision trees, each trained on different parts of the data using random samples and features. When making a prediction, the random forest takes the majority vote or averages the results from all the trees. This approach reduces overfitting, improves accuracy.

RESULT

1. Logged into my IBM Cloud Account

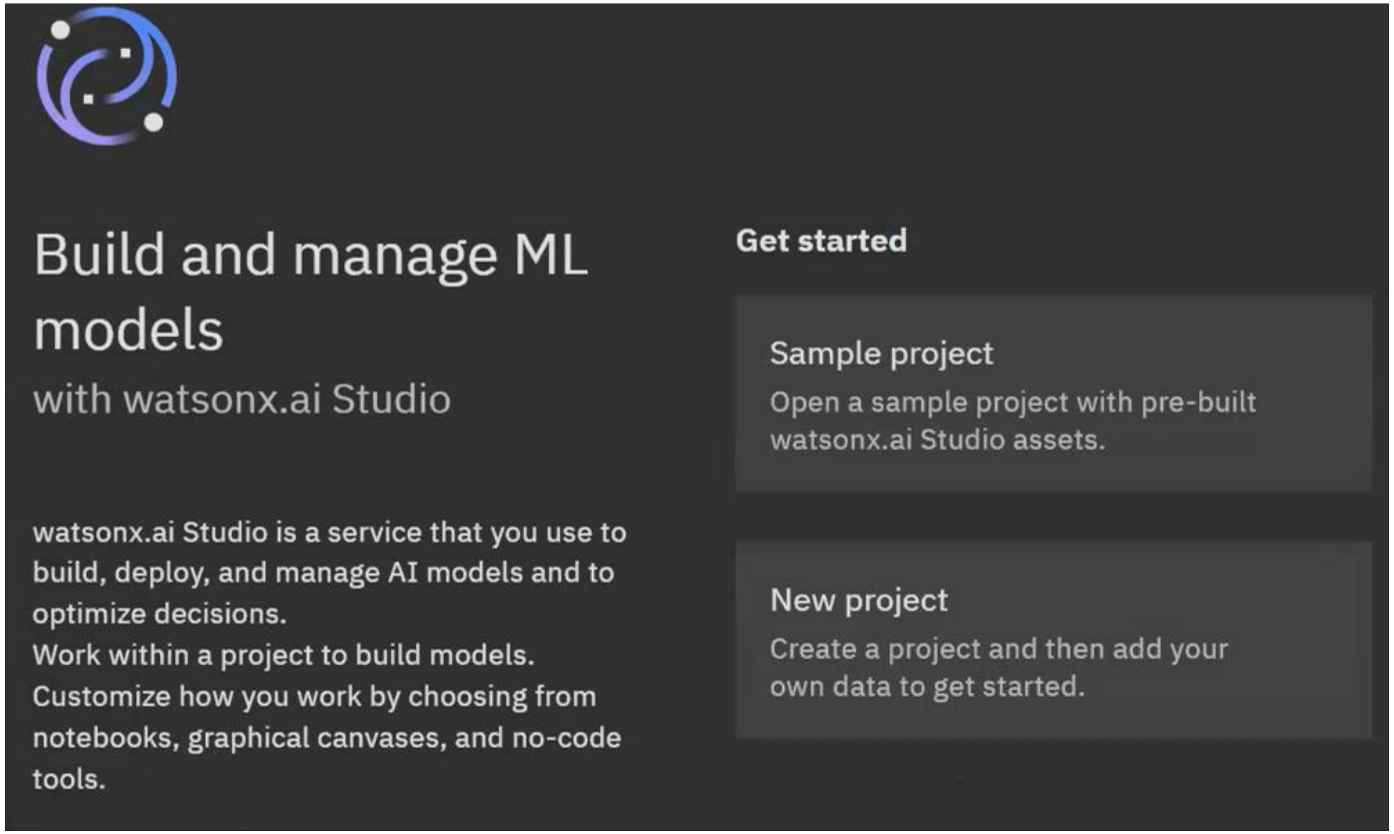
2. Re-created the required resources i.e Cloud Storage Object, Watsonx.ai Runtime and Watsonx.ai Studio

3. Created a New Project as Predictive_Maintenance_Project

4. Added the csv file **"predictive_maintenance.csv"** taken from Kaggle

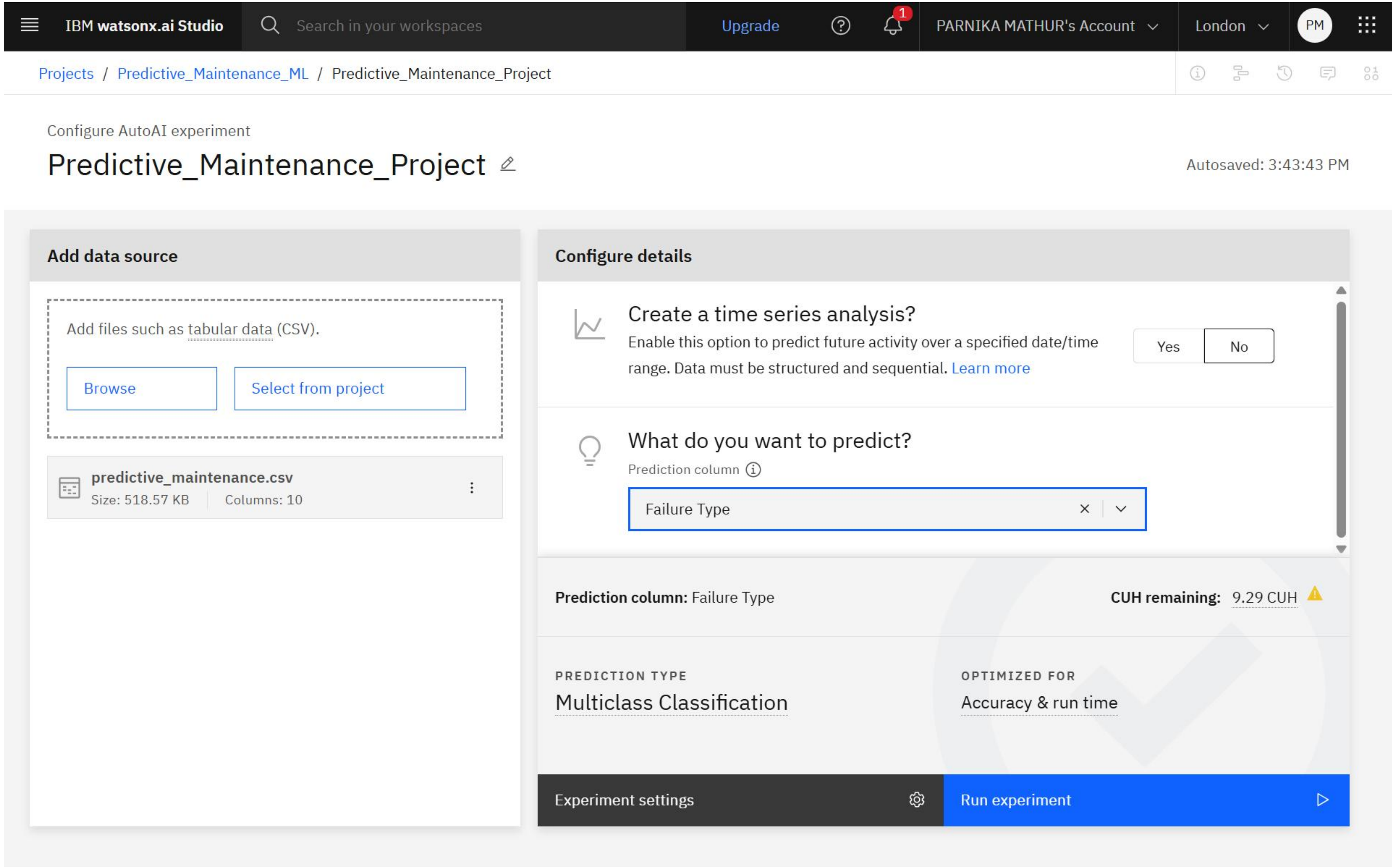
5. Data did not have ant time series analysis so selected the **Failure Type** column directly for classification prediction

6. Clicked on Run Experiment. (Slide 8)



edunet
foundation

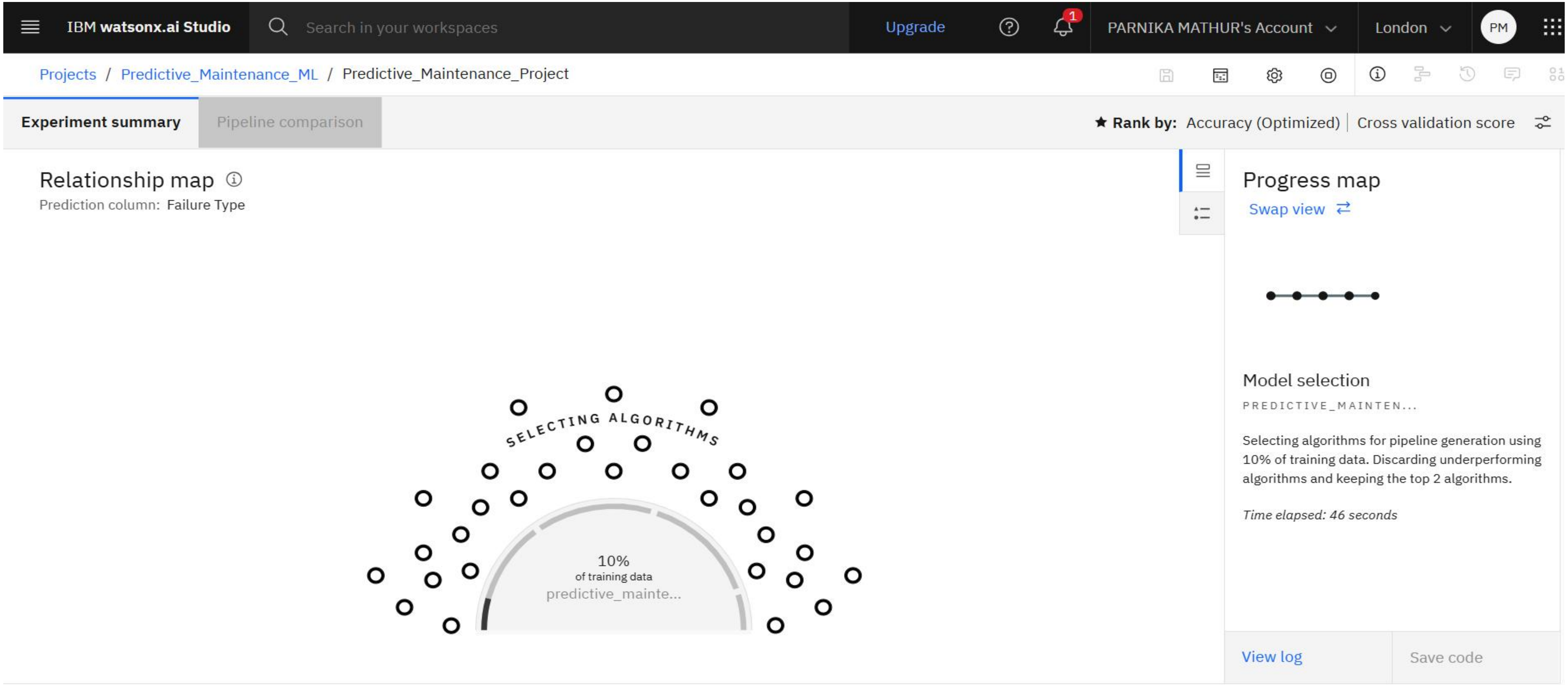
RESULT


IBM watsonx.ai Studio
Search in your workspaces
Upgrade
PARNIKA MATHUR's Account
London
PM
Projects / Predictive_Maintenance_ML / Predictive_Maintenance_Project
Configure AutoAI experiment
Predictive_Maintenance_Project
Autosaved: 3:43:43 PM
Add data source
Add files such as tabular data (CSV).
Browse
Select from project
predictive_maintenance.csv
Size: 518.57 KB
Columns: 10
Configure details
Create a time series analysis?
Enable this option to predict future activity over a specified date/time range. Data must be structured and sequential. Learn more
Yes
No
What do you want to predict?
Prediction column
Failure Type
Prediction column: Failure Type
CUH remaining: 9.29 CUH
PREDICTION TYPE
Multiclass Classification
OPTIMIZED FOR
Accuracy & run time
Experiment settings
Run experiment

RESULT (UPCOMMING SLIDE SUMMARY)

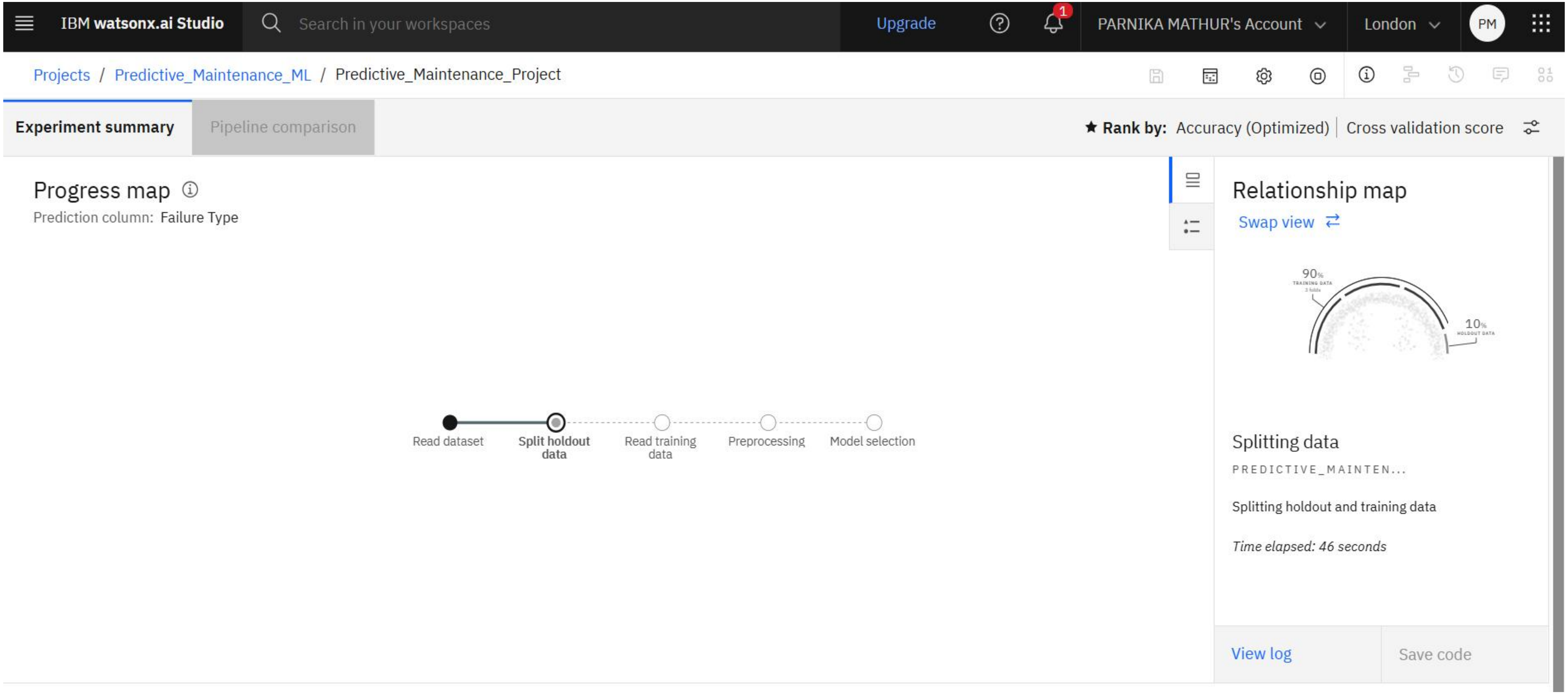
- **Slide 10 :** The IBM Cloud is automatic building relationships between the columns in the dataset and mapping them for algorithm processing and deciding overall algorithms to be used. Splitting the data 90% training and 10% testing.
- **Slide 11 :** Progress map showing the overall ML Pipeline for this Multiclassification Problem Statement.
- **Slide 12 :** Started building and testing multiple ML Pipelines using different algorithms.
- **Slide 13 :** Progress Map Relationship Map showing the internal working of ML Pipeline

RESULT



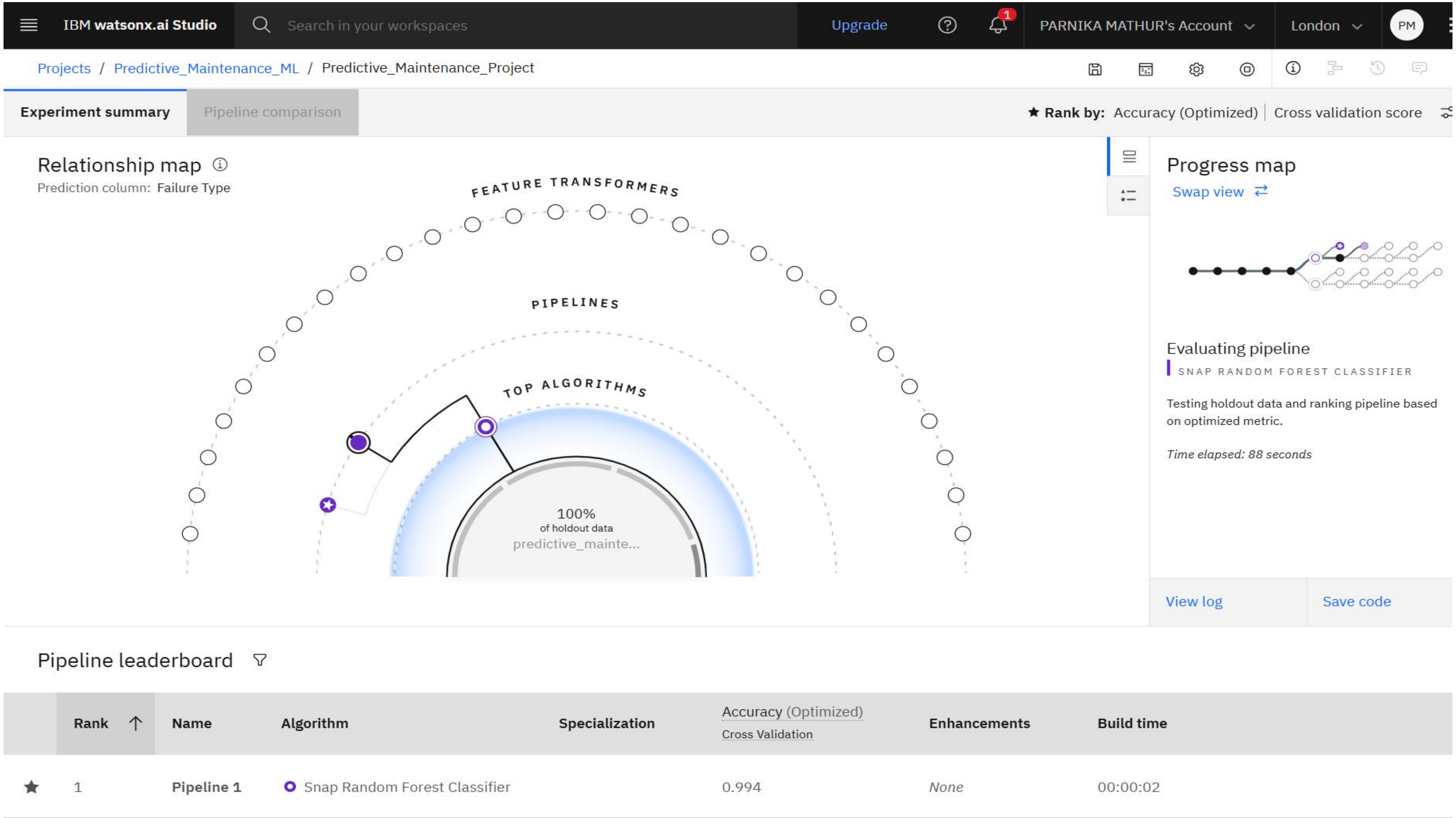
IBM watsonx.ai Studio
Search in your workspaces
Upgrade
1
PARNIKA MATHUR's Account
London
PM
Projects / Predictive_Maintenance_ML / Predictive_Maintenance_Project
Experiment summary
Pipeline comparison
Rank by: Accuracy (Optimized)
Cross validation score
Relationship map
Prediction column: Failure Type
SELECTING ALGORITHMS
10%
of training data
predictive_mainte...
Progress map
Swap view
Model selection
PREDICTIVE_MAINTEN...
Selecting algorithms for pipeline generation using 10% of training data. Discarding underperforming algorithms and keeping the top 2 algorithms.
Time elapsed: 46 seconds
View log
Save code

RESULT



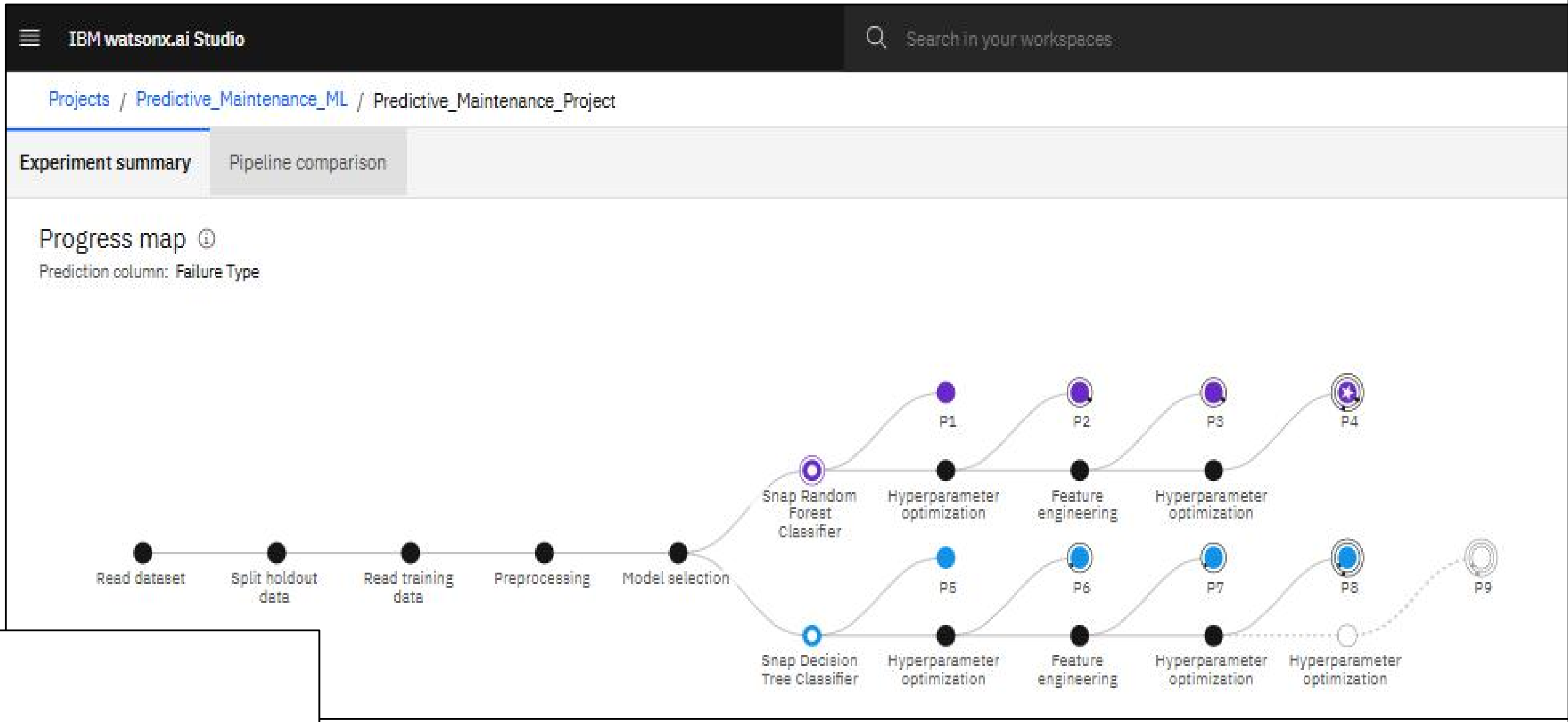
edunet foundation

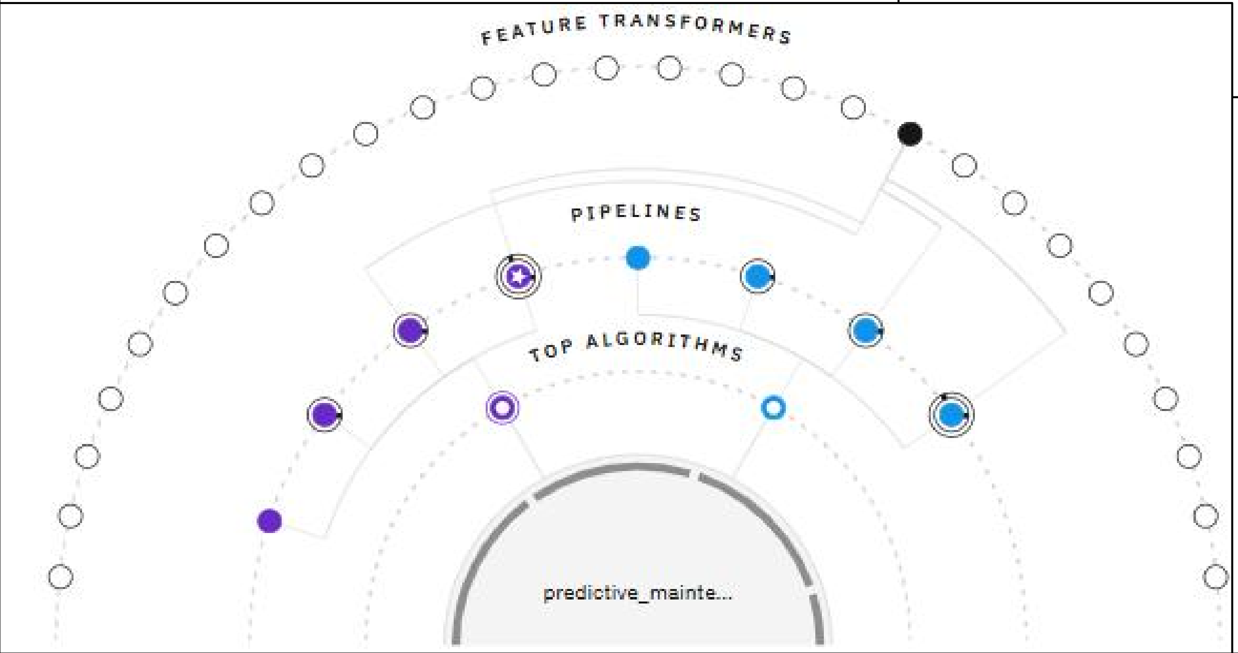
RESULT



edunet foundation

RESULT


IBM watsonx.ai Studio
Search in your workspaces
Projects / Predictive_Maintenance_ML / Predictive_Maintenance_Project
Experiment summary
Pipeline comparison
Progress map
Prediction column: Failure Type
P1
P2
P3
P4
Snap Random Forest Classifier
Hyperparameter optimization
Feature engineering
Hyperparameter optimization
Read dataset
Split holdout data
Read training data
Preprocessing
Model selection
P5
P6
P7
P8
P9
Snap Decision Tree Classifier
Hyperparameter optimization
Feature engineering
Hyperparameter optimization
Hyperparameter optimization


FEATURE TRANSFORMERS
PIPELINES
TOP ALGORITHMS
predictive_mainte...

RESULT

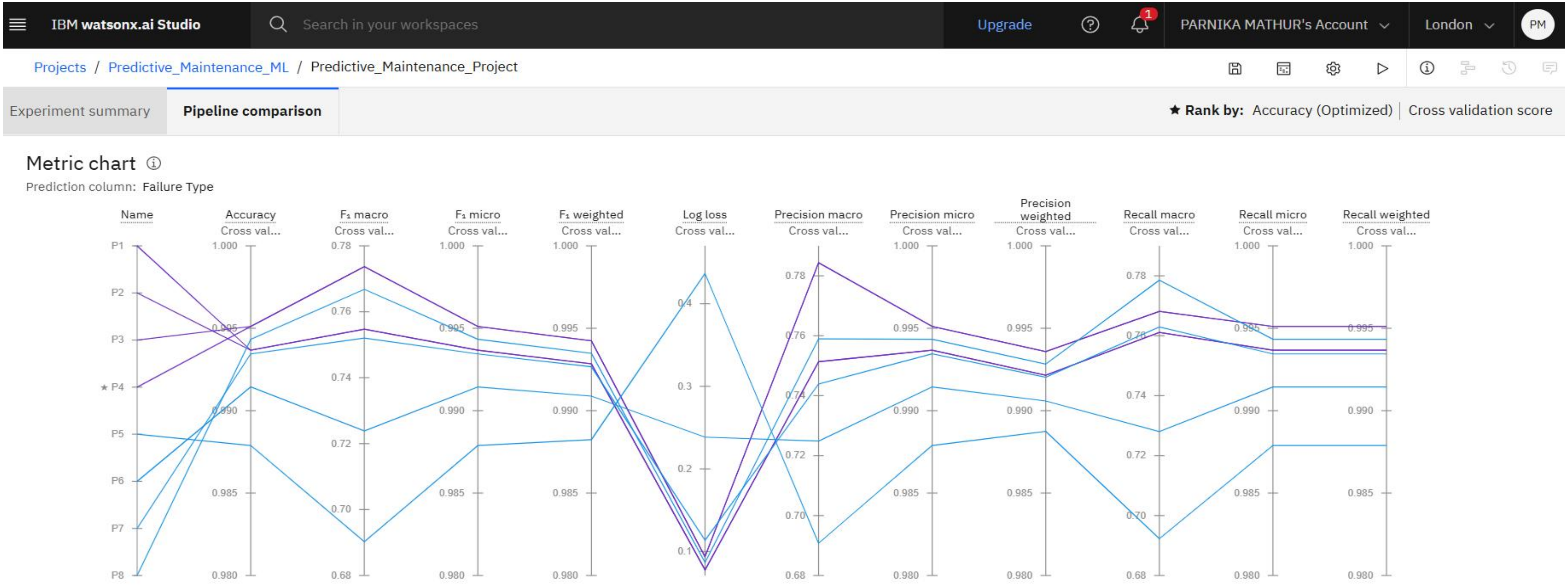
Final ML Pipeline created which displays all the algorithms tried and tested by IBM automation feature engineering hyperparameter tuning. (Snap Decision Tree Classifier and Snap Random Forest Classifier)

Pipeline leaderboard

	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 4	Snap Random Forest Classifier		0.995	HPO-1 FE HPO-2	00:00:39
	2	Pipeline 3	Snap Random Forest Classifier		0.995	HPO-1 FE	00:00:30
	3	Pipeline 8	Snap Decision Tree Classifier		0.994	HPO-1 FE HPO-2	00:00:26
	4	Pipeline 2	Snap Random Forest Classifier		0.994	HPO-1	00:00:06
	5	Pipeline 1	Snap Random Forest Classifier		0.994	*None*	00:00:02
	6	Pipeline 7	Snap Decision Tree Classifier		0.993	HPO-1 FE	00:00:23
	7	Pipeline 6	Snap Decision Tree Classifier		0.991	HPO-1	00:00:04
	8	Pipeline 5	Snap Decision Tree Classifier		0.988	*None*	00:00:01

RESULT

Metric Chart comparing different pipelines showing accuracy, F1, precision, recall, log loss and saving the model with maximum accuracy i.e P4 : Snap Random Forest Classifier



RESULT

Viewing the datatypes of all the features before deploying into proper IBM online space.

Projects / Predictive_Maintenance_ML / P4 - Snap Random Forest Classifier: Predictive_Maintenance_Project

Input (1)

Column	Type
Air temperature [K]	double
Process temperature [K]	double
Product ID	other
Rotational speed [rpm]	double
Target	double
Tool wear [min]	double
Torque [Nm]	double
Type	other

IBM watsonx.ai Studio   Search in your workspaces   Upgrade   1   PARNIKA MATHUR's Accou...   London   PM

Deployment spaces / Predictive_Maintenance / P4 - Snap Random Forest Classifier: Predictive_Maintenance_Project

Deployments   Model details

Search   New deployment

Name	Type	Status	Tags	Last modified
Predictive_Maintenance_Deployment	Online	Deployed		22 seconds ago PARNIKA MATHUR (You)

About this asset

Name

P4 - Snap Random Forest Classifier: Predictive_Maintenance_Project

Description

No description provided.

edunet foundation

RESULT

Input testing inputs for 9 feature set to verify for prediction accuracy and fault type

IBM watsonx.ai Studio | Search in your workspaces | Upgrade | 2 | PARNIKA MATHUR's Accou... | London | PM

Deployment spaces / Predictive_Maintenance / P4 - Snap Random Forest Classifier: Predictive_Maintenance_Project /

Predictive_Maintenance_Deployment

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

	UDI (double)	Product ID (other)	Type (other)	Air temperature [K] (double)	Process temperature [K] (double)	Rotational speed [rpm] (double)	Torque [Nm] (double)	Tool wear [min] (double)	Target (double)
1	40	L47219	L	298.8	309.1	1350	52.5	111	0
2	7456	H29425	H	259	300	1856	73.5	70	1
3	8922	M15056	M	302.5	305.2	2768	68.1	15	1
4	1597	L48775	L	290	307.4	1365	52.9	218	1
5	3963	M18795	M	302.6	311.6	1226	68.2	187	1
6	1749	H31126	H	298.4	307.7	1626	30.2	166	0
7	1222	M16081	M	297	308.3	1399	46.4	132	0
8	3237	M18096	M	300.7	310	1342	62.4	20	1
9	2865	H32278	H	300.5	309.4	1380	47.6	246	1

edunet foundation

RESULT

Prediction results

Display format for prediction results

(●) Table view ( ) JSON view

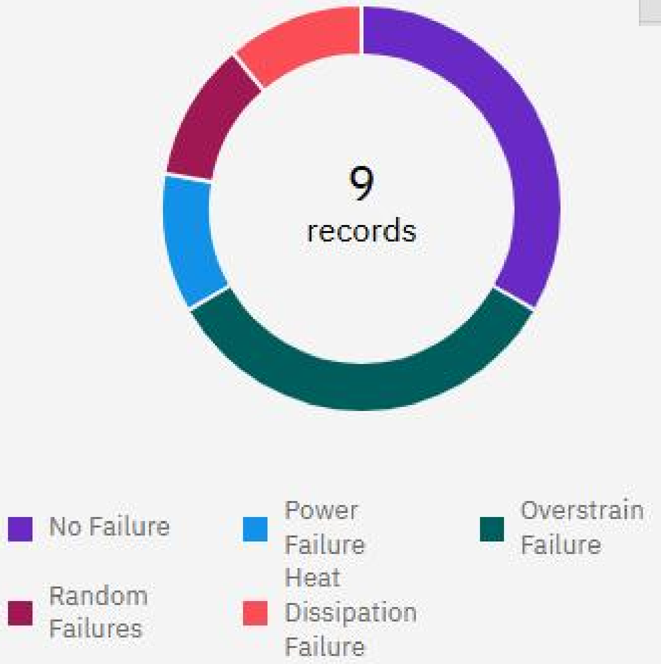
Show input data

	Prediction	Confidence
1	No Failure	100%
2	Power Failure	90%
3	No Failure	50%
4	Overstrain Failure	90%
5	Overstrain Failure	67%
6	No Failure	80%
7	Random Failures	60%
8	Heat Dissipation Failure	90%
9	Overstrain Failure	40%

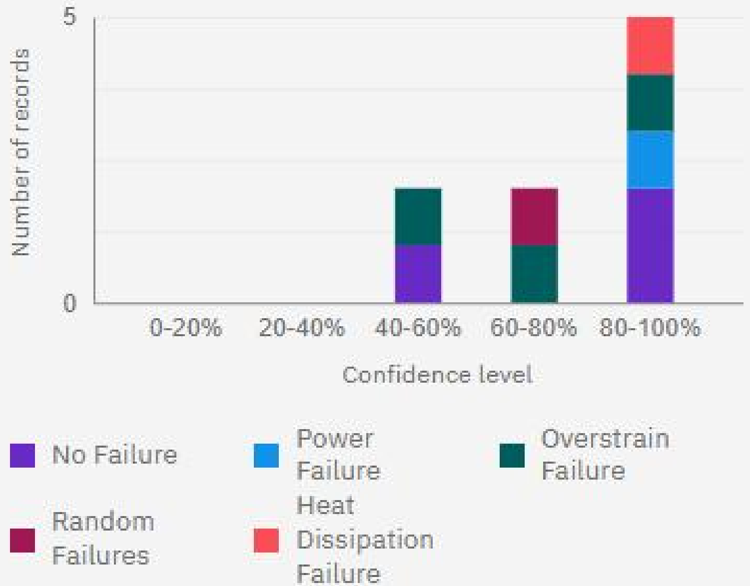
Prediction type

Multiclass classification

Prediction percentage



Confidence level distribution



RESULT

GitHub Repository Link :

https://github.com/ParnikaMathur15/IBM_Cloud_Predictive_Maintenance

My Deployed Model on IBM Cloud :

https://eu-gb.ml.cloud.ibm.com/ml/v4/deployments/e3445557-ec96-4290-900c-7356cbe52cbf/predictions?version=2021-05-01

CONCLUSION

- The predictive maintenance system developed using IBM Cloud AutoAI demonstrated strong performance with an overall accuracy of 99.5%, successfully identifying different types of machine failures from real-time sensor data. This highlights the effectiveness of automated model selection and deployment in industrial applications.
- However, despite the high accuracy, the average model confidence was around 40%, suggesting that while predictions were correct, the certainty behind them was relatively low. This could be due to overlapping sensor patterns between failure types or limited feature variability.
- The project showcases the potential of machine learning to reduce unplanned downtime, optimize maintenance schedules, and support smarter industrial operations. Moving forward, improvements like incorporating more contextual data, refining class boundaries, or applying ensemble techniques could help enhance model confidence and decision reliability in real-world use.
- I found out by directly putting the values from the csv to the prediction input and the model at times gave the right predicted value and in some an incorrect predicted value

FUTURE SCOPE

In the future, this Predictive Maintenance Model can be improved by handling more types of machine failures as more precise real-time data becomes available. Using advanced deep learning techniques like LSTM can help detect complex patterns in sensor data. By combining **edge computing** with the model, predictions can be made locally on devices, reducing latency and dependency on central servers. The system can also be extended to include **augmented reality (AR)** interfaces for technicians, showing real-time failure insights visually during inspections. Furthermore, integrating Web3 technologies logging of sensor data and maintenance history, making audits can be more reliable. his will help industries reduce unexpected breakdowns, improve machine life, and save maintenance costs.

REFERENCES

- **Kaggle Dataset :**

https://www.kaggle.com/datasets/shivamb/machine_x0002_predictive-maintenance-classification

- **Watsonx.AI Studio Documentation**

- **Research Papers :**

  - **A Machine Learning Framework for Predictive Maintenance in Industry 4.0 by** Matteo Galar, J. M. Fernández, E. Barrenechea, H. Bustince, F. Herrera, 2019
  - **Multiclass Failure Prediction of Industrial Machines Using Sensor Data and Deep Learning** by L. Wang, M. Azam, J. Zhang, H. Liang, 2021
  - **Machine Learning Methods for Predictive Maintenance of Industrial Equipment: A Survey** by Ahmed M. Eldeeb, Mohamed M. Abdelaty, et al., 2021

IBM CERTIFICATIONS (GETTING STARTED WITH AI)

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

PARNIKA MATHUR

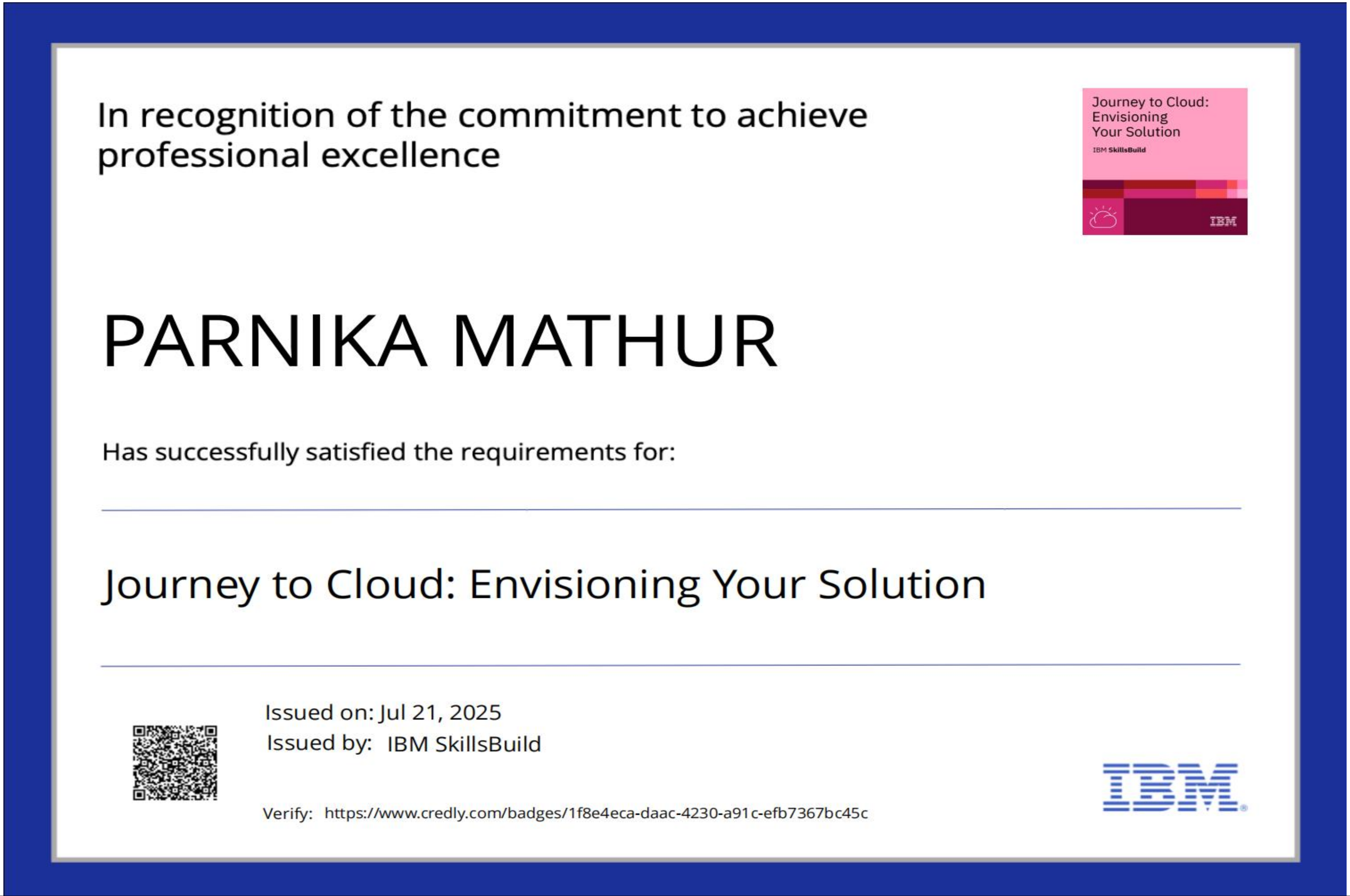
Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 20, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/2080e7e8-9e7b-4d2c-ada8-d28dc89c29ba

IBM CERTIFICATIONS (JOURNEY TO CLOUD)


In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM SkillsBuild

PARNIKA MATHUR

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/1f8e4eca-daac-4230-a91c-efb7367bc45c

IBM

edunet foundation

IBM CERTIFICATIONS (RAG LAB)

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Parnika Mathur

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 26 Jul 2025 (GMT)

Learning hours: 20 mins

THANK YOU

BY - PARNIKA MATHUR