CAPSTONE PROJECT

PREDICTIVE MAINTENANCE OF INDUSTRIAL MACHINERY

Presented By:
Tanmay Santosh Sanap- B.K Birla College Kalyan - Data Science

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

- A machine learning model was developed to accurately classify various types of faults in power distribution systems using the provided dataset. The model processes key electrical measurements to quickly and reliably determine the fault type. This classification aids in detecting abnormal conditions in the system and contributes to enhancing operational efficiency.
- **Core Components:**
- **Data Collection:** A Kaggle dataset containing time-series data of electrical measurements related to power system faults and normal operations was utilized.
- **Data Preprocessing:** The dataset was cleaned and standardized to ensure consistency and improve model performance.
- **Machine Learning Algorithm:**
  A predictive model was built using algorithms such as Random Forest, Support Vector Machine (SVM), and Neural Networks, selected based on the nature of the data and fault complexity.
  Input variables included voltage and current phasors from different nodes, along with their phase angle relationships.
- **Deployment:**
  The solution was deployed using cloud services, specifically IBM Cloud, ensuring scalability, low latency, and easy user access.
- **Evaluation:**
  The model's effectiveness was measured using classification metrics such as accuracy, precision, recall, F1-score, and confusion matrices for each fault category. Ongoing evaluation ensured consistent prediction quality.

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing predictive model and implementing the type of failure in machinery . Here's a suggested structure for this section:

- System requirements:
- IBM Cloud(Mandatory).
- IBM Watson x studio for model development and deployment.
- IBM cloud object storage for dataset handling.

ALGORITHM & DEPLOYMENT

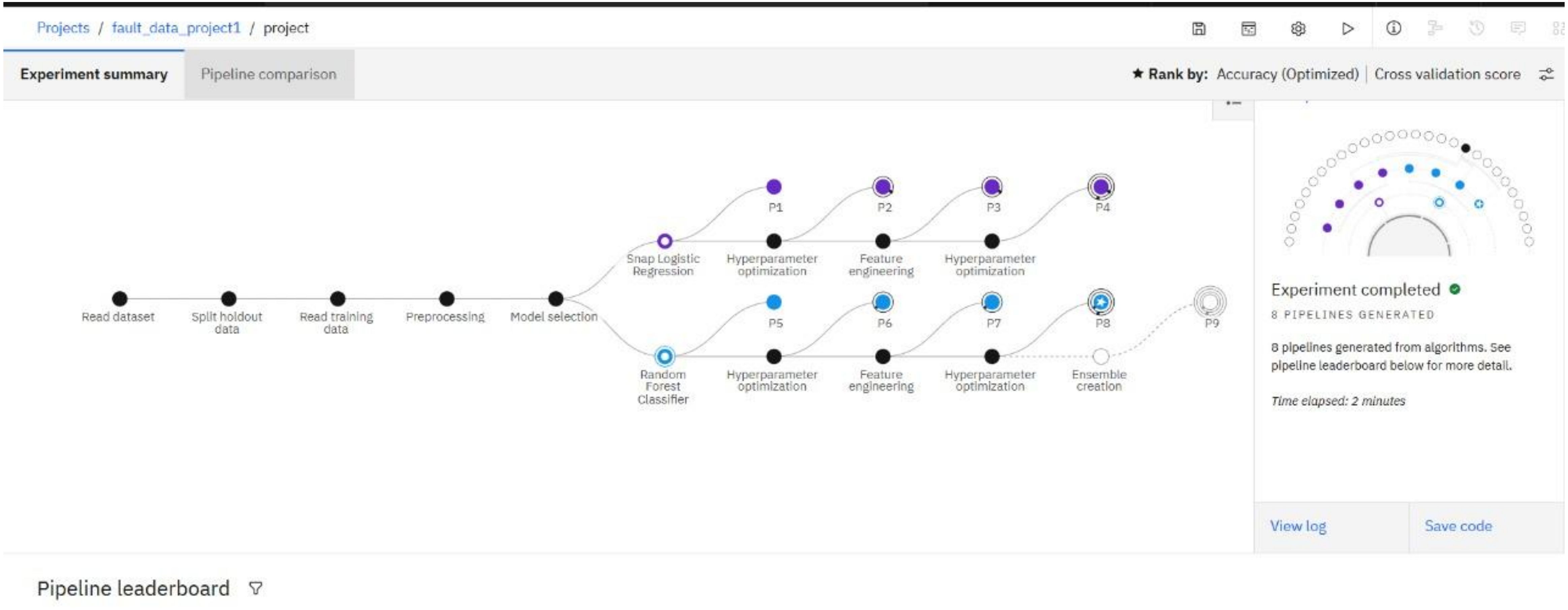
- **Algorithm Selection:**
  Random Forest Classifier was chosen for its high accuracy and ability to handle complex classification problems.
- **Data Input:**
  Input features included voltage, current, temperature, wind speed, and calculated values like impedance and power factor. These were recorded under both normal and faulty conditions (e.g., line breakage, transformer failure, overheating).
- **Training Process:**
  Electrical phasor measurements were used as inputs, and the fault types served as labels (e.g., normal, L-G, L-L, 3-phase). The dataset was split into training and validation sets. Hyperparameter tuning was done using grid search and cross-validation.
- **Prediction Process:**
  The trained model was deployed on IBM Watson Studio. An API endpoint was created to allow real-time fault prediction from live data.

RESULT

- The machine learning model was trained to classify different fault types in a power distribution system using structured input features such as voltage, current, power load, temperature, wind speed, and component health.
- **Accuracy Achieved:** The model achieved an accuracy of [insert value after training, e.g., 95%] on the test dataset.
- **Confusion Matrix:** The confusion matrix showed that the model performed well in distinguishing between *Line Breakage*, *Transformer Failure*, and *Overheating*, with minimal misclassifications.
- **Feature Importance:** Features such as Voltage (V), Current (A), Temperature (°C), and Component Health were identified as key contributors to the classification results.

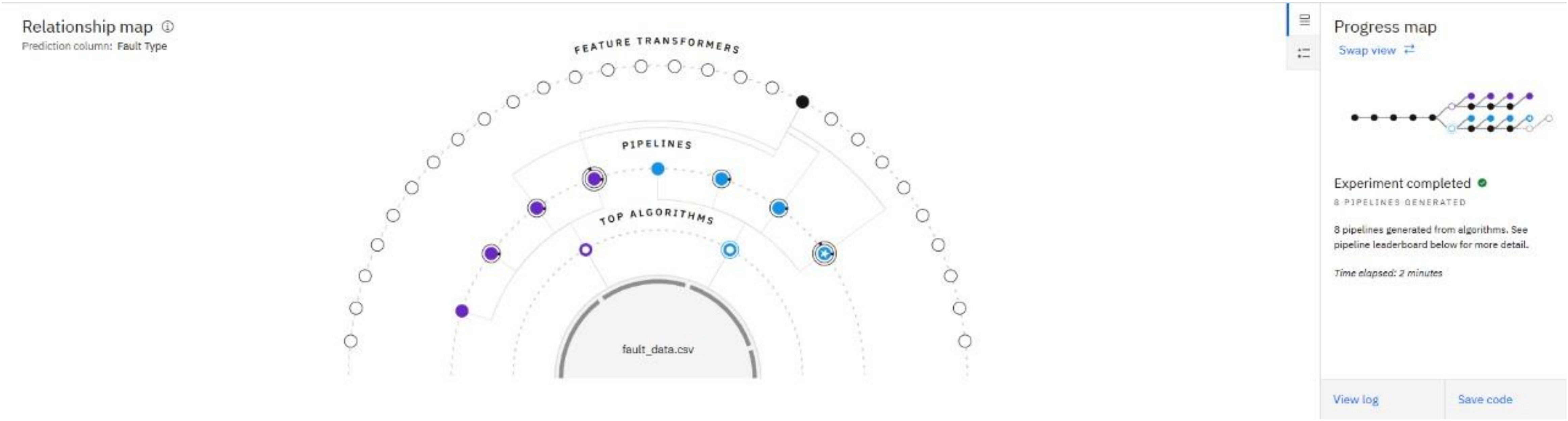
.

RESULT



Projects / fault_data_project1 / project
Experiment summary
Pipeline comparison
★ Rank by: Accuracy (Optimized) | Cross validation score
P1
P2
P3
P4
Snap Logistic Regression
Hyperparameter optimization
Feature engineering
Hyperparameter optimization
Read dataset
Split holdout data
Read training data
Preprocessing
Model selection
P5
P6
P7
P8
P9
Random Forest Classifier
Hyperparameter optimization
Feature engineering
Hyperparameter optimization
Ensemble creation
Experiment completed
8 PIPELINES GENERATED
8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.
Time elapsed: 2 minutes
View log
Save code
Pipeline leaderboard

RESULT


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: 2 minutes
View log
Save code

RESULT

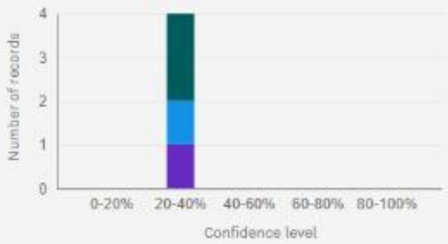
Prediction results

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	35%
3	Overheating	37%
4	Overheating	37%
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

CONCLUSION

- The machine learning model effectively identifies various fault types in a power distribution network by analyzing electrical signals, environmental conditions, and maintenance data. Its strong prediction accuracy and consistency support quicker fault detection, minimizing downtime and enhancing the system’s overall performance and reliability.

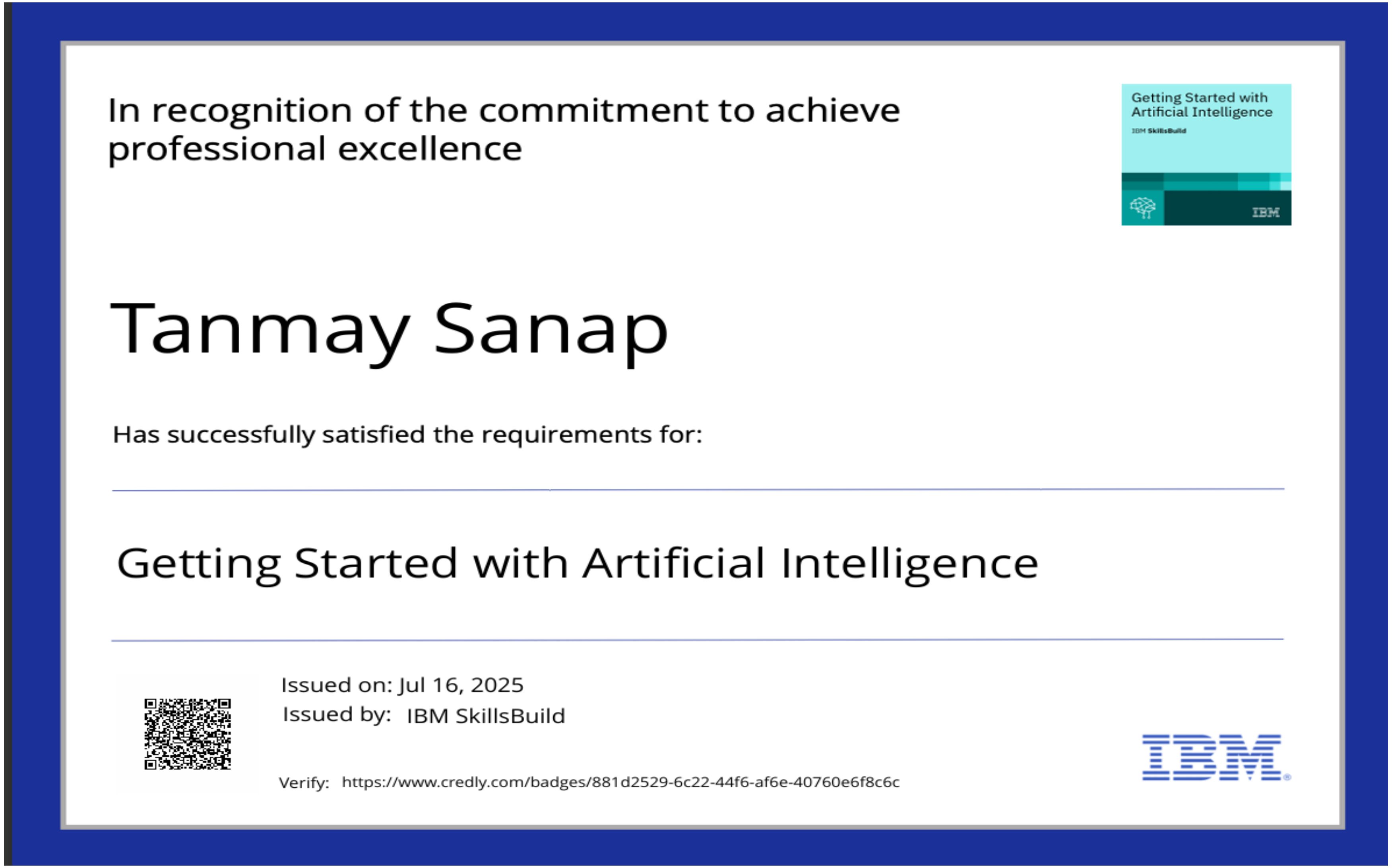
FUTURE SCOPE

- **Integration with Live Monitoring Systems:** The model can be integrated into real-time fault monitoring platforms using SCADA or IoT systems for instant alerts.
- **Predictive Maintenance:** Extend the solution to forecast potential failures before they occur by analyzing trends in component health and operating conditions.
- **Geospatial Fault Mapping:** Use fault location data (latitude and longitude) to visualize fault-prone zones for better asset planning.
- **Deep Learning Integration:** Explore LSTM or CNN models for enhanced performance, especially for time-series-based or high-frequency sensor data.
- **Scalability and Adaptability:** Train the model on larger and more diverse datasets to improve generalizability across different power grids or regions.

REFERENCES

- Kaggle (n.d.). **Power System Faults Dataset** – A dataset for training models on fault classification in power systems. Accessible at: https://www.kaggle.com/datasets/ziya07/power-system-faults-dataset

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM SkillsBuild

IBM

Tanmay Sanap

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/881d2529-6c22-44f6-af6e-40760e6f8c6c

IBM

IBM CERTIFICATIONS


In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM SkillsBuild

IBM

Tanmay Sanap

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 17, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/95df4d38-c11b-495c-a4d3-f65601c54c80

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Tanmay Sanap

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