CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:

VARSHA SHIVANAND K – RV UNIVERSITY – B.Tech CSE (AI/ML Major)

GitHub Repo Link : https://github.com/varshashivanand-k/Capstone-Project--Intelligent-Classification-of-Rural-Infrastructure-Projects.git

edunet foundation

OUTLINE :

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

PROBLEM STATEMENT :

- The Pradhan Mantri Gram Sadak Yojana (PMGSY) is a flagship rural development program in India, initiated to provide all-weather road connectivity to eligible unconnected habitations. Over the years, the program has evolved through different phases or schemes (PMGSY-I, PMGSY-II, RCPLWEA, etc.), each with potentially distinct objectives, funding mechanisms, and project specifications. For government bodies, infrastructure planners, and policy analysts, efficiently categorizing thousands of ongoing and completed projects is crucial for effective monitoring, transparent budget allocation, and assessing the long-term impact of these schemes. Manual classification is time-consuming, prone to errors, and scales poorly.

PROPOSED SOLUTION :

- The proposed system aims to **classify rural infrastructure projects** into their respective **PMGSY schemes** (such as PMGSY-I, PMGSY-II, RCPLWEA, etc.) based on **physical** and **financial characteristics**. This will support government bodies in **efficient monitoring**, **budgeting,** and **analysis**.
- **Data Collection:**
  - Fetch the **Pradhan Mantri Gram Sadak Yojana (PMGSY)** Dataset from the AI Kosh Portal.
  - AI Kosh dataset link – https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_y ojna_pmgsy.html
- **Data Preprocessing:**
  - Perform basic cleaning like removing irrelevant unnamed columns.
- **Machine Learning Algorithm:**
  - Explore various machine learning algorithms such as: **XGB Classifier**, **Snap Random Forest Classifier,** etc.
  - Use **AutoAI** in **IBM Watson Studio** to automate model selection, feature engineering, and hyperparameter tuning.
  - Select the best pipeline based on accuracy.
- **Deployment:**
  - Deploy the solution on a scalable and reliable platform such as **IBM Cloud**.
- **Result:**
  - Display the predicted **PMGSY Scheme** and confidence scores.
  - Allow users to download or export predictions for planning/reporting.

SYSTEM APPROACH :

The **“System Approach”** outlines the cloud-based workflow used to develop and deploy a classification model for identifying the correct PMGSY scheme based on physical and financial project characteristics.

1. **System Requirements**
    - **Platform**: IBM Cloud (Watson Studio)
    - **Environment**: AutoAI experiment environment
    - **No local installation** is required as all processing, training, and evaluation are conducted in the cloud.

2. Libraries & Tools Used

The following Python libraries are implicitly used by IBM AutoAI during the modeling phase:

- **pandas**, **numpy** – For data manipulation and internal processing.
- **scikit-learn** – For preprocessing, metrics, and evaluation pipelines.
- **xgboost** – Selected machine learning algorithm (XGBoost Classifier).
- **IBM SnapML** – Considered in other pipelines, but not used in the final selected model.

ALGORITHM & DEPLOYMENT :

- **Algorithm Selection :**
  - AutoAI generated multiple pipelines for comparison.
  - **Pipeline 8 (XGB)** achieved the highest accuracy **(92.4%).**
  - It outperformed others due to better preprocessing and tuning The leaderboard summary is shown in the next slide.
  - Hence, Pipeline 8 was selected and deployed for prediction.

- **Data Input :**
  - Input features include physical and financial attributes of each project.
  - **Key Groups:** Location (State, District), Sanction Details (roads, bridges, cost), Completion Info, Financial Info (Expenditure), and remaining work.
  - These factors influence classification under the PMGSY Scheme.
  - **Target Variable**: PMGSY_SCHEME.

AUTOAI PIPELINE LEADERBOARD AND FINAL MODEL SELECTION

Pipeline leaderboard

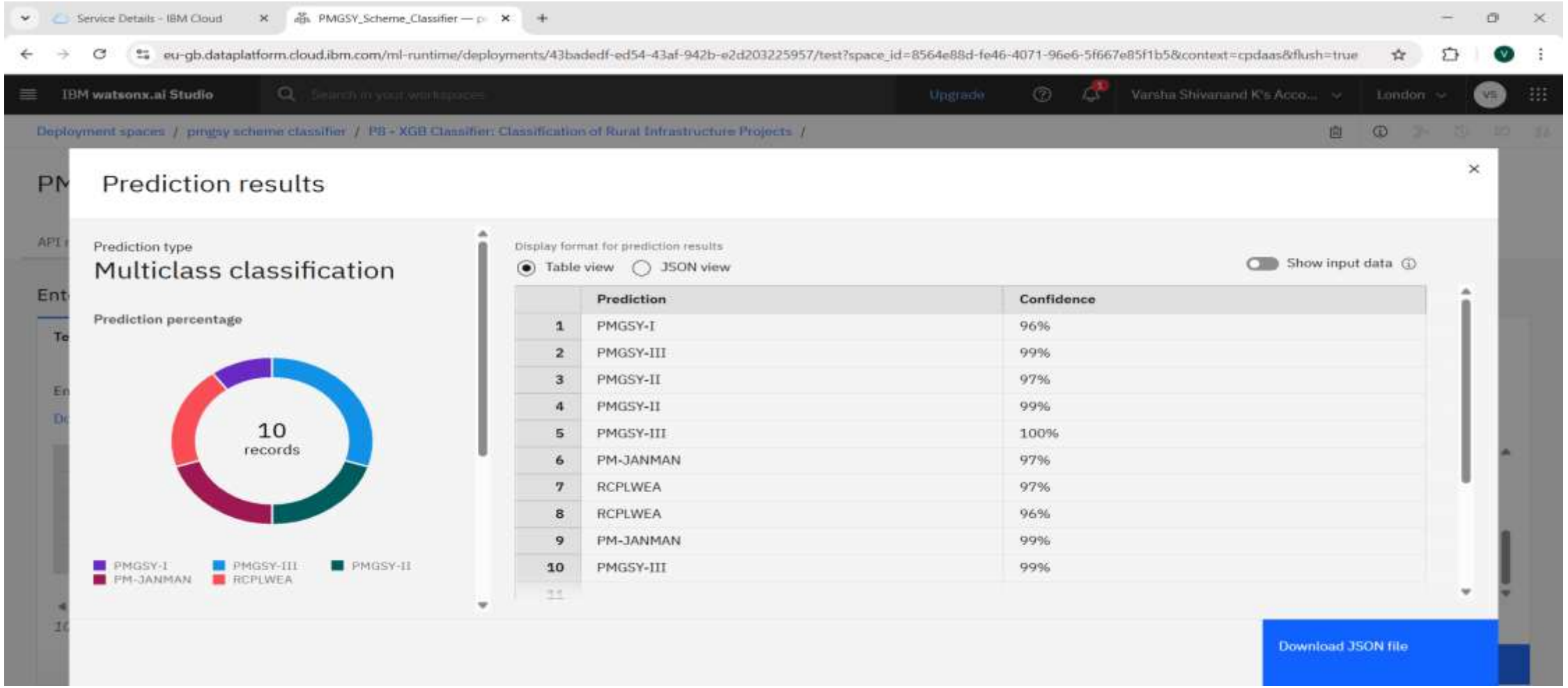
	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time	
★	1	Pipeline 8	XGB Classifier		0.924	HPO-1 FE HPO-2	00:01:58	Save as
	2	Pipeline 7	XGB Classifier		0.924	HPO-1 FE	00:01:15	
	3	Pipeline 6	XGB Classifier		0.918	HPO-1	00:00:24	
	4	Pipeline 5	XGB Classifier		0.918	None	00:00:03	
	5	Pipeline 4	Snap Random Forest Classifier		0.899	HPO-1 FE HPO-2	00:00:36	
	6	Pipeline 3	Snap Random Forest Classifier		0.899	HPO-1 FE	00:00:28	
	7	Pipeline 2	Snap Random Forest Classifier		0.897	HPO-1	00:00:07	
	8	Pipeline 1	Snap Random Forest Classifier		0.897	None	00:00:02	

ALGORITHM & DEPLOYMENT :

- **Training Process :**
  - The training process was handled by **IBM AutoAI**, which automatically performed key steps such as encoding categorical variables, performing feature engineering, and optimizing hyperparameters.
  - The final model was selected based on its leaderboard accuracy and generalization capability.

- **Prediction Process :**
  - After training, the final **XGBoost Classifier** Model was used to make predictions on new data inputs that were entered manually. It is also possible to upload a .CSV file for giving inputs.
  - Each input had the same structure and formatting as the training data, ensuring consistency.
  - The model processed the data and returned predicted classifications for the **PMGSY_SCHEME** variable.
  - Although the current setup uses manual input, the model can be deployed in a real-time environment or integrated into dashboards for automatic classification of new project records.

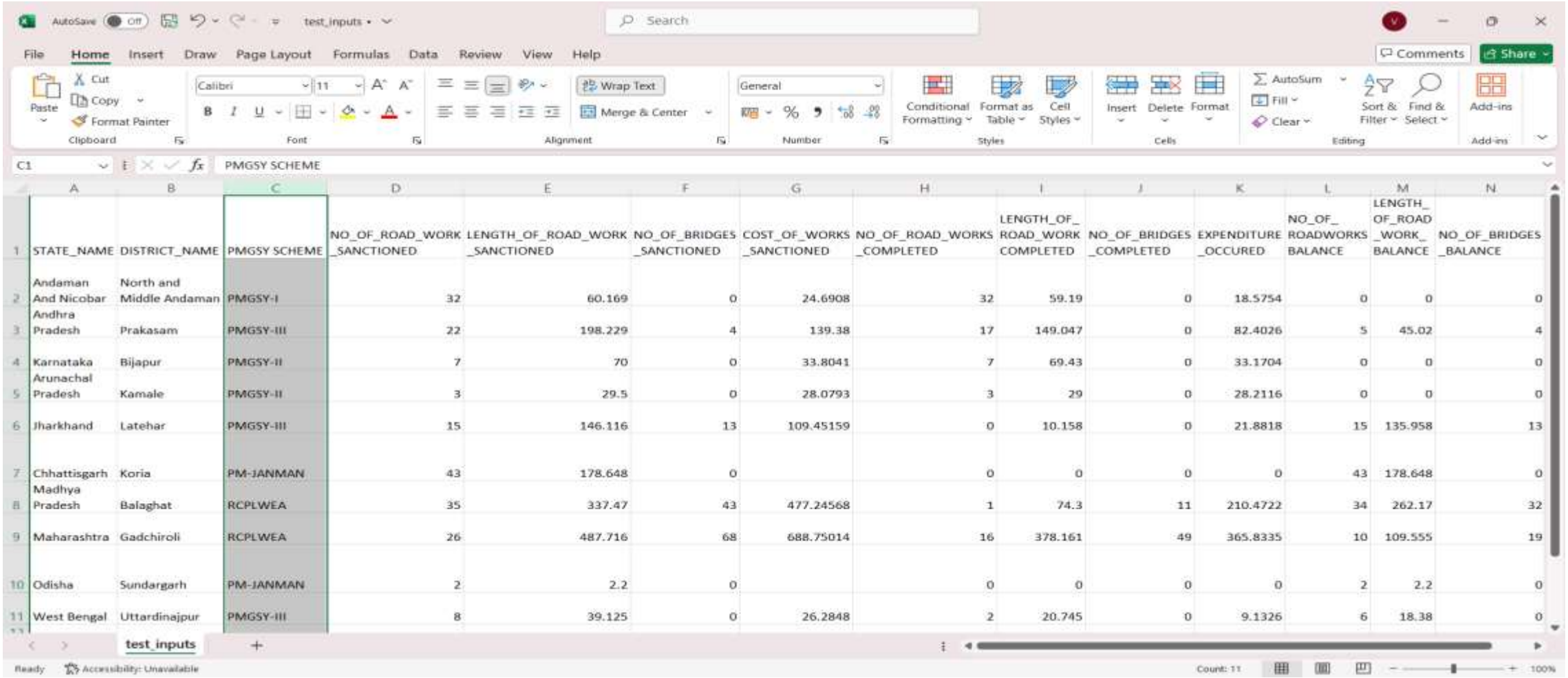
RESULT :



These are the PMGSY Scheme prediction results for the 10 input records, along with the confidence percentages.
The model demonstrates strong reliability with confidence scores ranging from **96% to 100%**.
The input data is provided in the next slide.

edunet foundation

INPUT DATA :



	STATE_NAME	DISTRICT_NAME	PMGSY SCHEME	NO_OF_ROAD_WORK_SANCTIONED	LENGTH_OF_ROAD_WORK_SANCTIONED	NO_OF_BRIDGES_SANCTIONED	COST_OF_WORKS_SANCTIONED	NO_OF_ROAD_WORKS_COMPLETED	LENGTH_OF_ROAD_WORK COMPLETED	NO_OF_BRIDGES_COMPLETED	EXPENDITURE_OCCURED	NO_OF_ROADWORKS BALANCE	LENGTH_OF_ROAD_WORK_BALANCE	NO_OF_BRIDGES_BALANCE
2	Andaman And Nicobar	North and Middle Andaman	PMGSY-I	32	60.169	0	24.6908	32	59.19	0	18.5754	0	0	0
3	Andhra Pradesh	Prakasam	PMGSY-III	22	198.229	4	139.38	17	149.047	0	82.4026	5	45.02	4
4	Karnataka	Bijapur	PMGSY-II	7	70	0	33.8041	7	69.43	0	33.1704	0	0	0
5	Arunachal Pradesh	Kamale	PMGSY-II	3	29.5	0	28.0793	3	29	0	28.2116	0	0	0
6	Jharkhand	Latehar	PMGSY-III	15	146.116	13	109.45159	0	10.158	0	21.8818	15	135.958	13
7	Chhattisgarh	Koria	PM-JANMAN	43	178.648	0		0	0	0	0	43	178.648	0
8	Madhya Pradesh	Balaghat	RCPLWEA	35	337.47	43	477.24568	1	74.3	11	210.4722	34	262.17	32
9	Maharashtra	Gadchiroli	RCPLWEA	26	487.716	68	688.75014	16	378.161	49	365.8335	10	109.555	19
10	Odisha	Sundargarh	PM-JANMAN	2	2.2	0		0	0	0	0	2	2.2	0
11	West Bengal	Uttardinajpur	PMGSY-III	8	39.125	0	26.2848	2	20.745	0	9.1326	6	18.38	0

Here are the input records used for prediction, including the actual PMGSY Scheme labels.
The actual PMGSY Scheme labels match the predicted values, confirming the model's **100% accuracy** on this sample.

edunet foundation

CONCLUSION :

- This project successfully demonstrates how machine learning, specifically **the XGB Classifier**, can be effectively used to classify rural infrastructure projects under the appropriate PMGSY scheme.
- The model achieved a test accuracy of **92.40%**, and in our predictions on unseen data, it **correctly classified 100% of the entries**, with **confidence scores ranging from 96% to 100%**, indicating high reliability.
- During implementation, challenges like missing or imbalanced data were addressed using **AutoAI's** built-in **feature engineering** and **hyperparameter tuning.**
- The results validate the usefulness of AI in improving **transparency**, **efficiency**, and **data-driven decision-making** in large-scale government programs.

FUTURE SCOPE :

- **More Data Sources:** Including real-time project updates, weather, terrain and socio-economic data to boost accuracy
- **Model Optimization:** Using advanced tuning or trying other models
- **Wider Scope:** Expanding to other rural infrastructure schemes.
- **Deployment:** Developing a user dashboard and API for easy government use

REFERENCES :

- **IBM AutoAI Documentation**

  IBM Cloud Pak for Data: AutoAI for machine learning

  https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/autoai-overview.html?context=cpdaas

- **IndiaAI - AI Kosh PMGSY Dataset**

  Pradhan Mantri Gram Sadak Yojana (PMGSY) Dataset, sourced from AI Kosh, IndiaAI.

  https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_y ojna_pmgsy.html

IBM CERTIFICATIONS :

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM

Varsha Shivanand K

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/aa220d78-f3b5-403e-a7ed-d5ff7687b49b

IBM

edunet foundation

IBM CERTIFICATIONS :

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM

Varsha Shivanand K

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

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

Verify: https://www.credly.com/badges/c501ad3c-14c0-41ab-a96e-f8e38ec5a179

IBM

edunet foundation

IBM CERTIFICATIONS :

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Varsha Shivanand K

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT) **Learning hours:** 20 mins

edunet foundation

THANK YOU

edunet
foundation