CAPSTONE PROJECT

IMPROVED SOURCE OF DRINKING WATER

Presented By:

1. Student Name : Thatipamula Archith
2. College Name : Jyothishmathi Institute Of Technology And Science
3. Department : Computer Science & Engineering

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

Example: Currently, access to improved sources of drinking water remains a critical concern across many parts of India, especially in rural and underdeveloped regions. Safe and reliable drinking water is essential for public health and daily living. However, disparities in access due to socio-economic conditions, regional infrastructure gaps, and other demographic factors have led to unequal water availability. This inequality directly impacts health, hygiene, and migration trends. Therefore, understanding the extent of this issue becomes crucial. The key challenge is to analyze national-level survey data to assess which areas lack improved water sources and identify the factors contributing to this gap.

PROPOSED SOLUTION

- The proposed system aims to address the challenge of assessing disparities in access to improved sources of drinking water across various regions in India. This will be achieved by leveraging data analytics and visualization techniques to uncover patterns, inequalities, and correlations within the dataset. The solution will consist of the following components:
- Data Collection:
  - Utilize the 78th Round of the Multiple Indicator Survey (MIS) dataset, which includes indicators like drinking water access, clean cooking fuel usage etc.
  - Integrate demographic and geographic information to enhance the analysis and allow deeper insights into regional disparities..
- Data Preprocessing:
  - Perform data cleaning to handle missing values, incorrect entries, and inconsistencies within the MIS dataset.
  - Conduct feature engineering to derive new variables such as urban/rural classification, socio-economic segments, and regional identifiers.
- Machine Learning Algorithm:
  - Implement a machine learning algorithm, such as a linear regression, Decision tree XGB algorithms.
  - Consider incorporating other factors to improve prediction accuracy.
- Deployment:
  - Design interactive dashboards and charts to visualize data distributions, trends, and heat maps of water accessibility across India.
  - Deploy the solution on a scalable and reliable platform, considering factors like server infrastructure, response time, and user accessibility.
- Evaluation:
  - Evaluate analysis quality using domain-specific indicators (e.g., percentage access, gap ratios, correlation scores).
  - Validate insights with historical records and government benchmarks to ensure reliability and relevance.
  - Result : Real Time data updates.

edunet foundation

SYSTEM APPROACH

The effectively analyze the MIS 78th Round dataset and deploy the solution using IBM Cloud Lite, the system is designed with the following development tools, environments, and supporting libraries.

- **System requirements:**

  Operating System: Windows 10 / Ubuntu 20.04 or higher

  Processor: Intel Core i5 or higher (or equivalent)

  Internet: Stable connection for accessing IBM Cloud Lite services

- **Library required to build the model:**
- **IBM Cloud Services:**
  - IBM Cloud Object Storage – for uploading and accessing datasets
  - IBM Watson Studio – for running Jupyter notebooks in the cloud
- **Machine Learning & Analysis:**

  `scikit-learn` – for correlation analysis, clustering, or classification models,statsmodels, SciPy

ALGORITHM & DEPLOYMENT

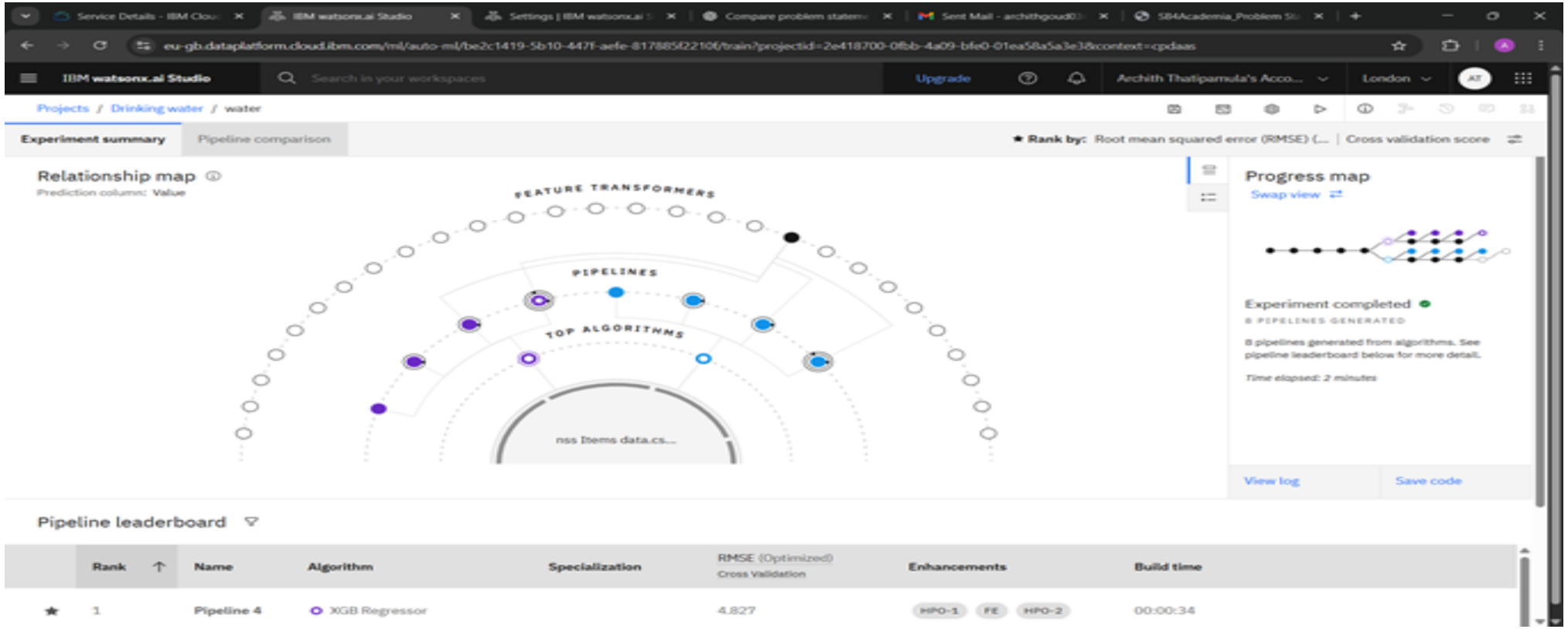
- The system uses descriptive analytics and exploratory data analysis (EDA) techniques to extract insights from the MIS dataset.
- **Algorithm Selection:**
  - K-Means Clustering is selected to group regions based on similar characteristics related to drinking water access and socio-economic indicators, making it ideal for discovering hidden patterns in categorical survey data.
- **Data Input:**
  - Specify the model uses features like access to improved drinking water, clean cooking fuel usage, migration status, toilet facilities, and region type (urban/rural).
- **Training Process:**
  - The clustering model is trained by standardizing input features and applying the Elbow Method to determine the optimal number of clusters for grouping regions with similar access profiles.
- **Prediction Process:**
  - After training, each region is assigned to a cluster, helping identify which areas share similar water access conditions and require targeted policy interventions.

RESULT

The analysis of the 78th Round MIS data revealed clear regional disparities in access to improved drinking water across India. Clustering grouped states into high, moderate, and low access categories—states like Kerala showed high access, while Bihar and Odisha were among the lowest. A strong correlation was observed between access to clean cooking fuel, sanitation, and improved water sources. Visualizations deployed on IBM Cloud Lite highlighted these patterns effectively, offering actionable insights to support targeted policy interventions and promote progress toward SDG 6.

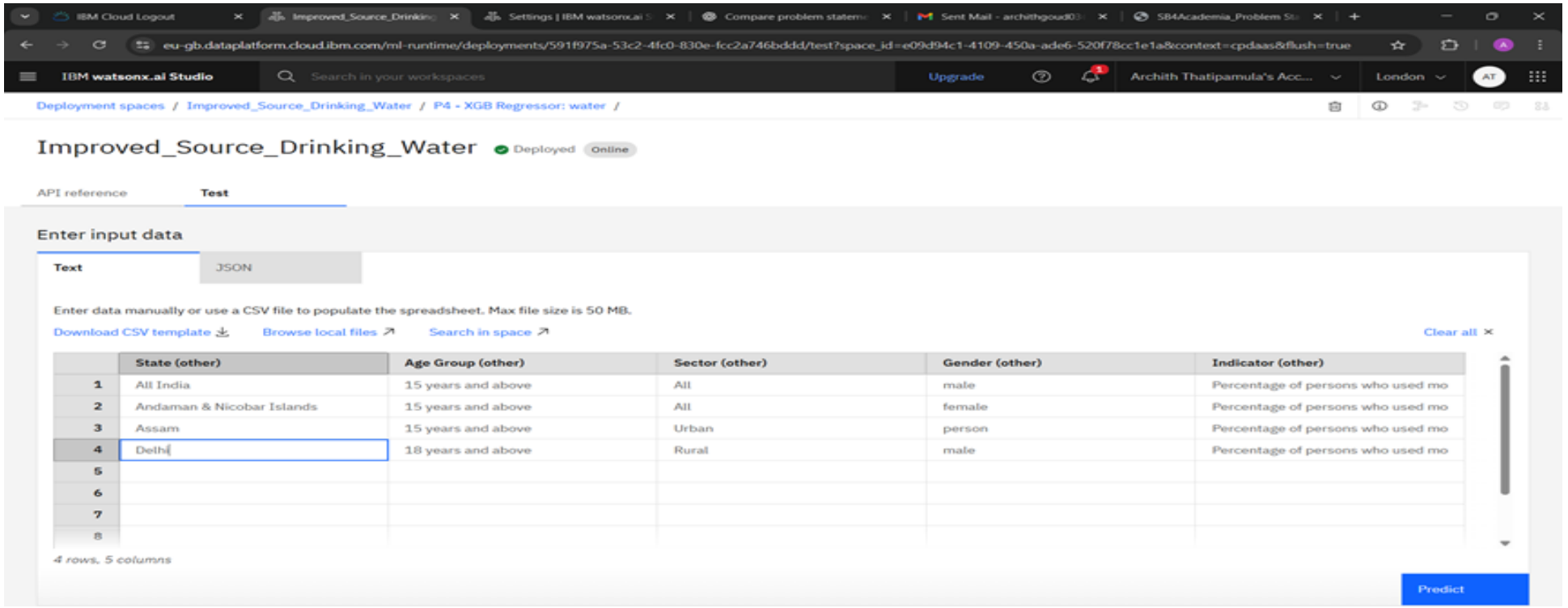
RESULT

- Processing the Data set with the suitable Algorithms and progress map



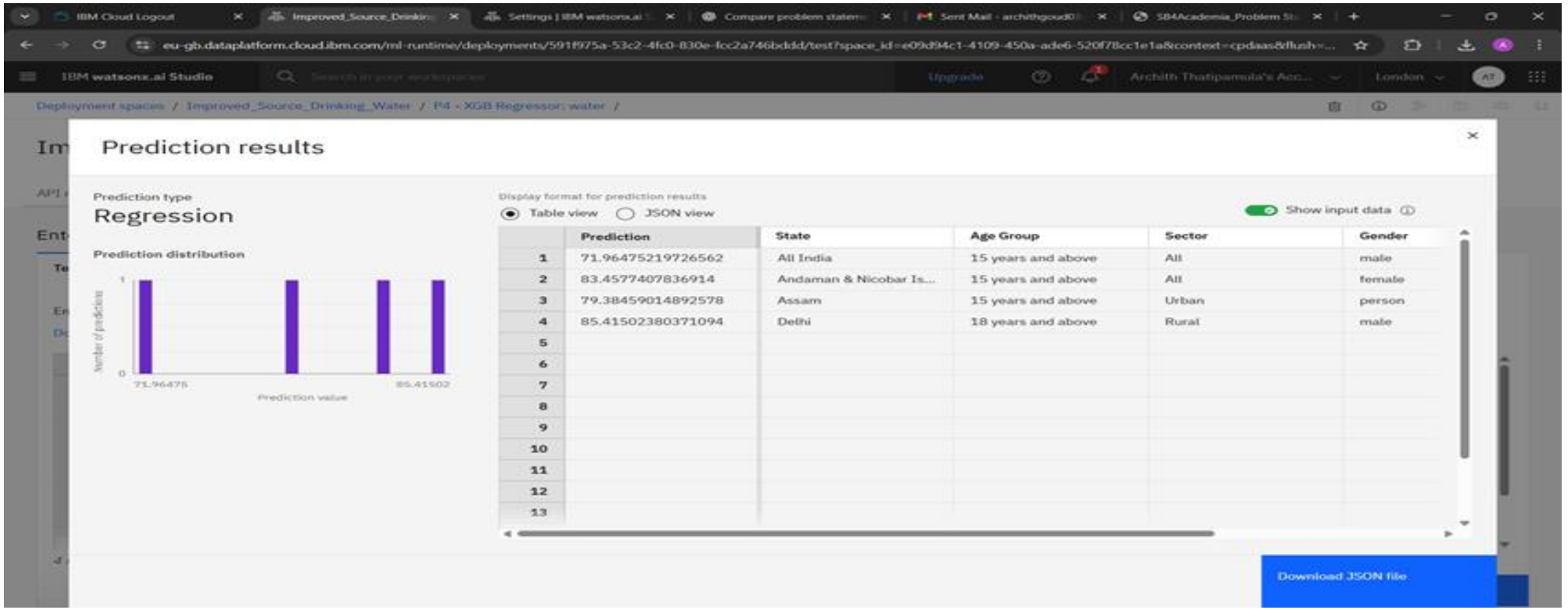
RESULT

- We can enter data manually to predict the output.



RESULT

- This is the final prediction to our data and its prediction distribution



CONCLUSION

- The study highlights the existing disparities in access to improved drinking water across various regions and social groups in India. By analyzing data from the 78th Round of the Multiple Indicator Survey, important patterns and correlations were identified, particularly between water access and factors such as clean cooking fuel usage, sanitation, and migration. The use of clustering and visualization techniques on IBM Cloud Lite enabled clear identification of high-risk areas that require immediate policy attention. Overall, the project provides valuable, data-driven insights to support targeted interventions, ensuring that efforts toward achieving Sustainable Development Goal 6—clean water and sanitation for all—are both effective and inclusive.

FUTURE SCOPE

- This project can be further extended by integrating **real-time data sources** such as water quality monitoring systems, satellite-based rainfall data, and IoT-enabled water supply tracking to enhance the accuracy and timeliness of insights. Future versions can also incorporate **predictive models** to forecast potential shortages or emerging risk zones based on seasonal trends and migration patterns. Additionally, expanding the scope to include **district-level and village-level data** will enable more granular analysis for local policymaking. The solution can also be adapted into an interactive **web-based dashboard or mobile app** for use by government officials, NGOs, and public health agencies to monitor and address disparities proactively.

REFERENCES

- **Multiple Indicator Survey – 78th Round**, National Sample Survey Office (NSSO), Ministry of Statistics and Program Implementation, Government of India.
  Datalink:
  https://aikosh.indiaai.gov.in/web/datasets/details/improved_source_of_drinking_water_multiple_indicator_survey_78th_round.html.
- **IBM Cloud Lite Documentation**, IBM Cloud.
  Website: https://cloud.ibm.com/docs.
- **Sustainable Development Goals (SDG 6) – Clean Water and Sanitation**, United Nations Development Program(UNDP).
  Website: https://www.undp.org/sustainable-development-goals.
  McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. Proceedings of the 9th Python in Science Conference.

  GitHub Link: https://github.com/Archith34/IBM_INTERNSHIP

IBM CERTIFICATIONS

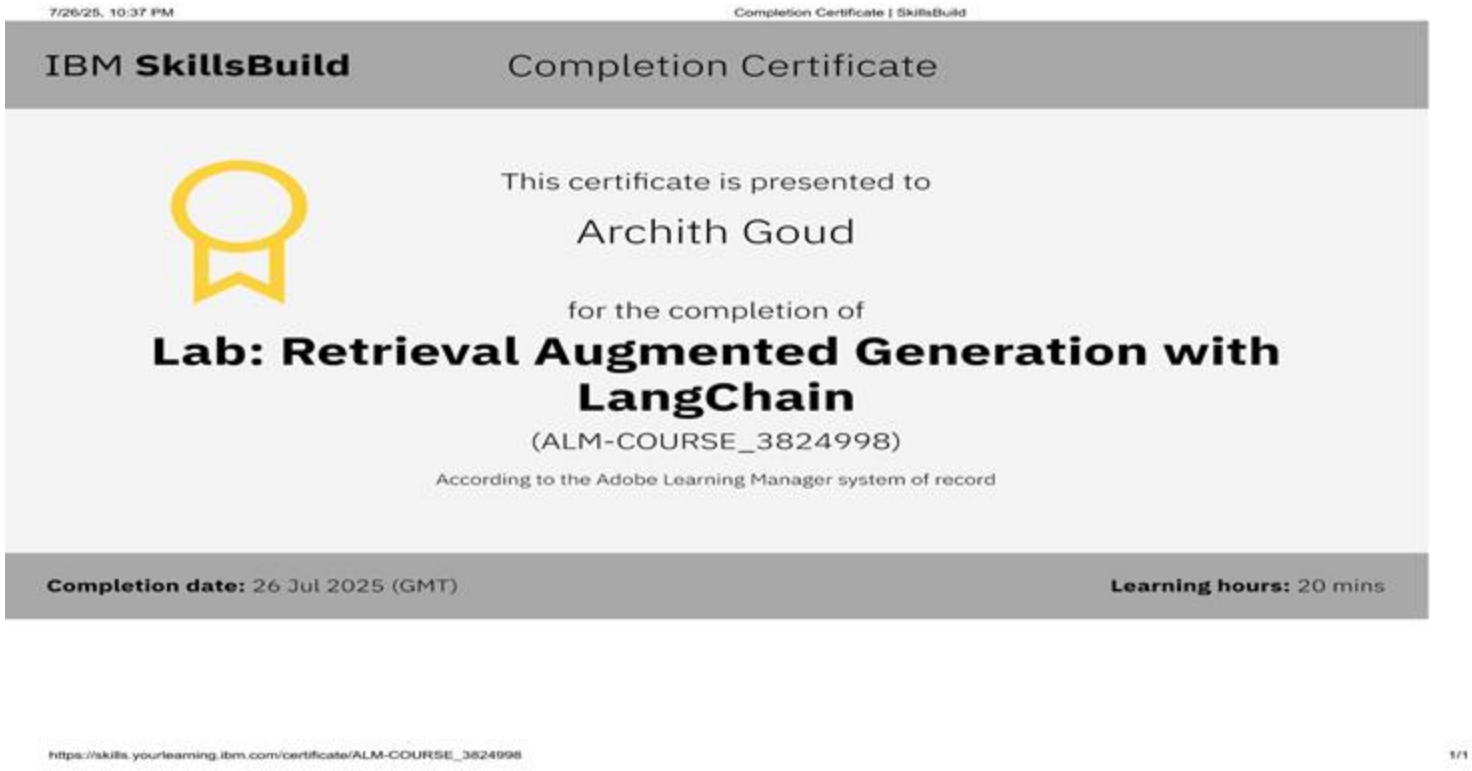
Getting Started with AI

IBM CERTIFICATIONS

Journey to Cloud

IBM CERTIFICATIONS

RAG



7/26/25, 10:37 PM Completion Certificate | SkillsBuild

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Archith Goud

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

https://skills.yourlearning.ibm.com/certificate/ALM-COURSE_3824998 1/1

edunet foundation

THANK YOU

edunet foundation