CAPSTONE PROJECT

IMPROVED SOURCE OF DRINKING WATER

Presented By:
Shankar Anil
JSS Academy of Technical Education, Noida, UP
Computer Science and Engineering(Data Science)

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

- Access to safe and improved sources of drinking water remains a critical challenge in India, particularly in rural and underdeveloped regions. Despite ongoing efforts under the Sustainable Development Goals (SDGs), significant inequalities in water accessibility persist across states and socio-economic groups. This study also aims to explore related indicators, including the use of clean cooking fuel and migration trends. The ultimate objective is to support equitable access to clean water and contribute meaningfully to India's progress on relevant SDG targets.

PROPOSED SOLUTION

- **Project Overview:**
  - **This project aims to analyze data from the 78th Round of the Multiple Indicator Survey (MIS) to assess the percentage of the population with access to improved drinking water sources.**
  - **By identifying patterns and disparities, the study aims to generate actionable insights to support evidence-based policymaking.**
- **Data Collection:**
  - **Collect historical data from the 78th Round of the Multiple Indicator Survey (MIS), including variables like State, Age, Gender, Sector (Urban/Rural), and relevant indicators.**
- **Data Preprocessing:**
  - **Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.**
  - **Feature engineering to extract relevant features from the data that might impact**
- **Machine Learning Algorithm:**
  - **Platform: IBM Watsonx.ai Studio (AutoAI)**
  - **Algorithm: Batched Tree Ensemble Regressor (XGBoost)RMSE: 4.827 (optimized using cross-validation)**
  - **Enhancements Used:HPO-1 & HPO-2 (Hyperparameter Optimization)FE (Feature Engineering)BATCH (Batched Training)**

PROPOSED SOLUTION

- **Deployment:**
  - **Assess the model's performance using appropriate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or other relevant metrics.**
  - **Fine-tune the model based on feedback and continuous monitoring of prediction accuracy.**
- **Result:**
  - **Developed and deployed a predictive model that estimates access to improved drinking water sources using inputs such as State, Age Group, Sector (Urban/Rural), Gender, and Indicator.**
- **Final Output:**
  - **A predictive model capable of forecasting water access using inputs like:**
  - **→ State, Age, Gender, Sector, and Key Indicators.**
  - **A visualization/dashboard tool for policymakers to analyze disparities and target interventions effectively.**

SYSTEM APPROACH

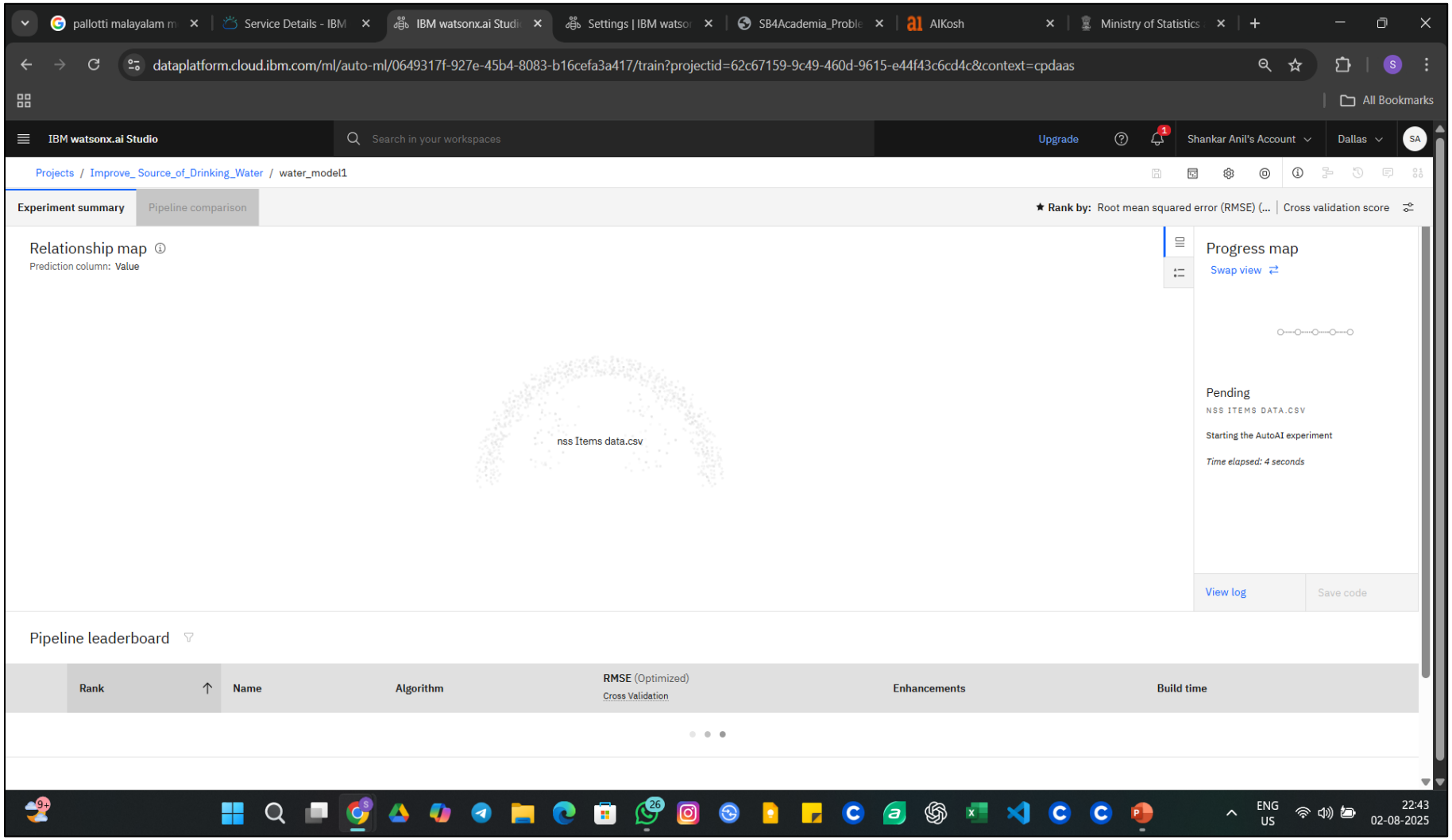
- **System requirements :-**

- **Processor: Intel Core i5 or higher**
- **RAM: Minimum 8 GB**
- **Operating System: Windows 10 / Linux / macOS**
- **Storage: At least 2 GB free space**
- **IBM Cloud Account**

ALGORITHM & DEPLOYMENT

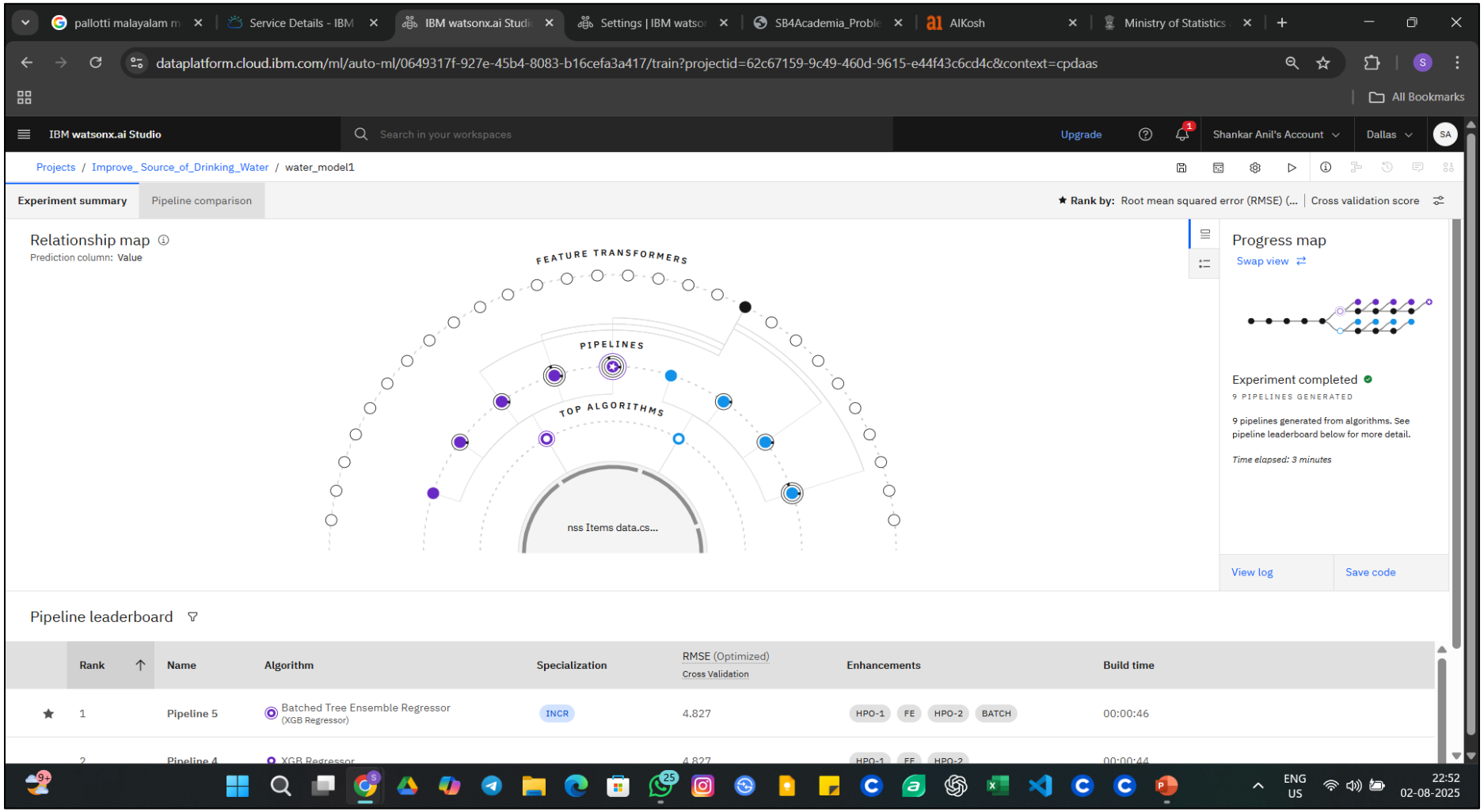
- **Algorithm Selection:**
  - **Model Chosen: Batched Tree Ensemble Regressor (XGBoost Regressor) selected via IBM Watsonx AutoAI based on performance metrics.**
- **Data Input:**
  - **Input Features: State, Age Group, Sector (Urban/Rural), Gender, and Indicator.**
- **Training Process:**
  - **Platform: IBM Watsonx AutoAI.**
  - **Training Phase P8: Utilized Random Forest Classifier (as part of AutoAI experimentation and comparison)..**
- **Prediction Process:**
  - **Final predictions generated using the selected Random Forest Classifier, based on the inputs: State, Age Group, Sector, Gender, Indicator.**

RESULT (SCREENSHOTS)



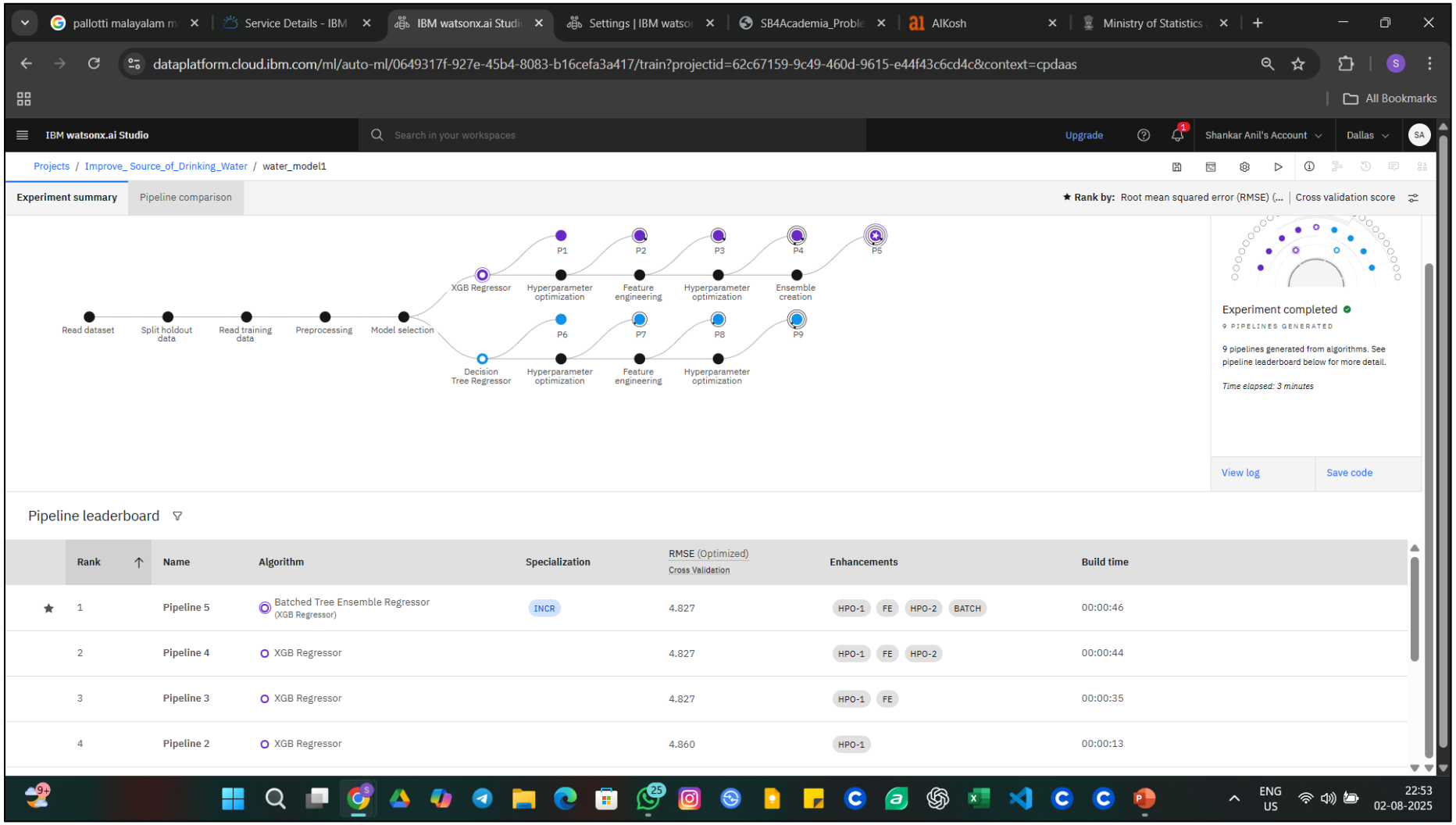
edunet foundation

RESULT (SCREENSHOTS)



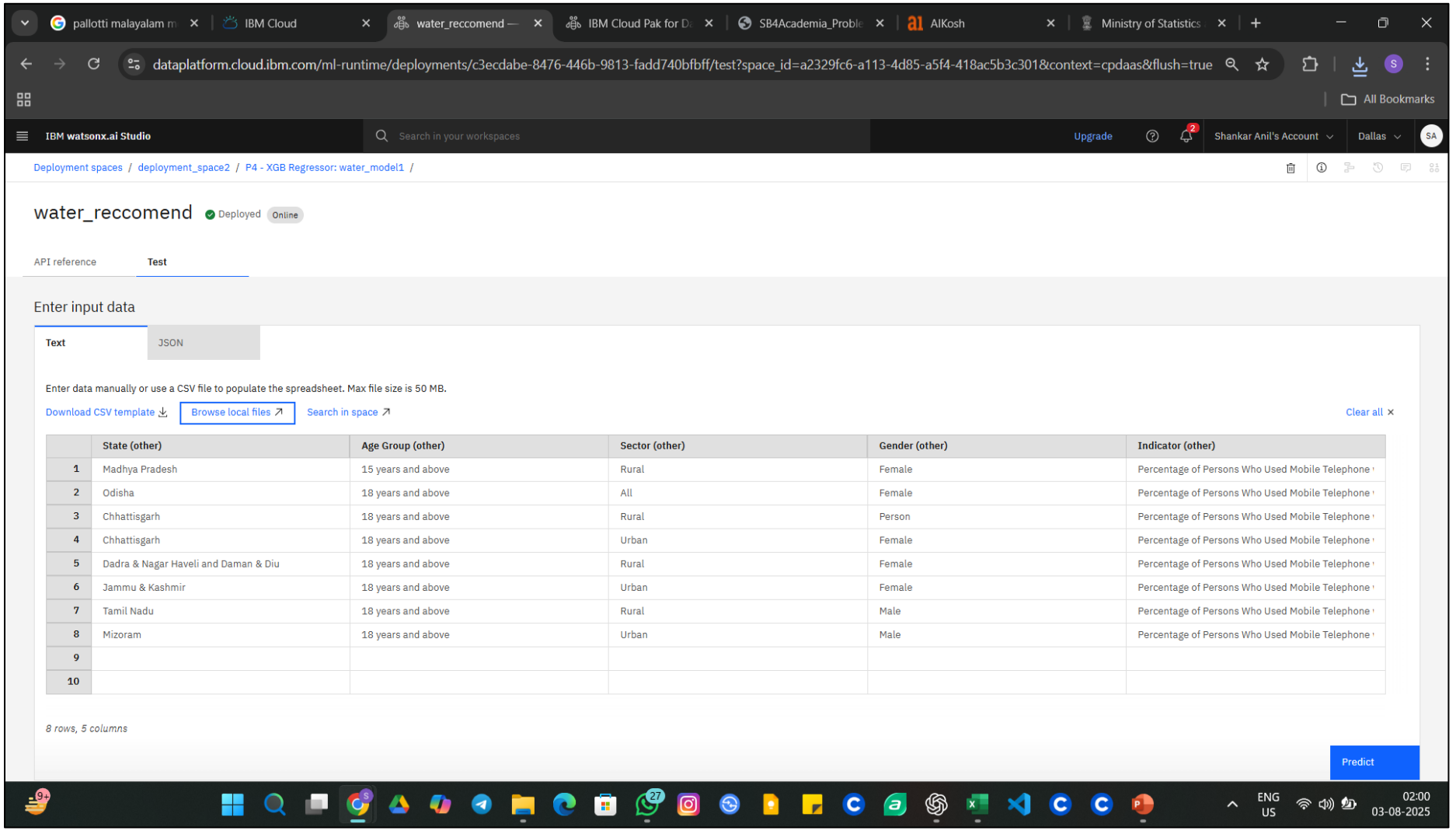
edunet foundation

RESULT (SCREENSHOTS)



edunet foundation

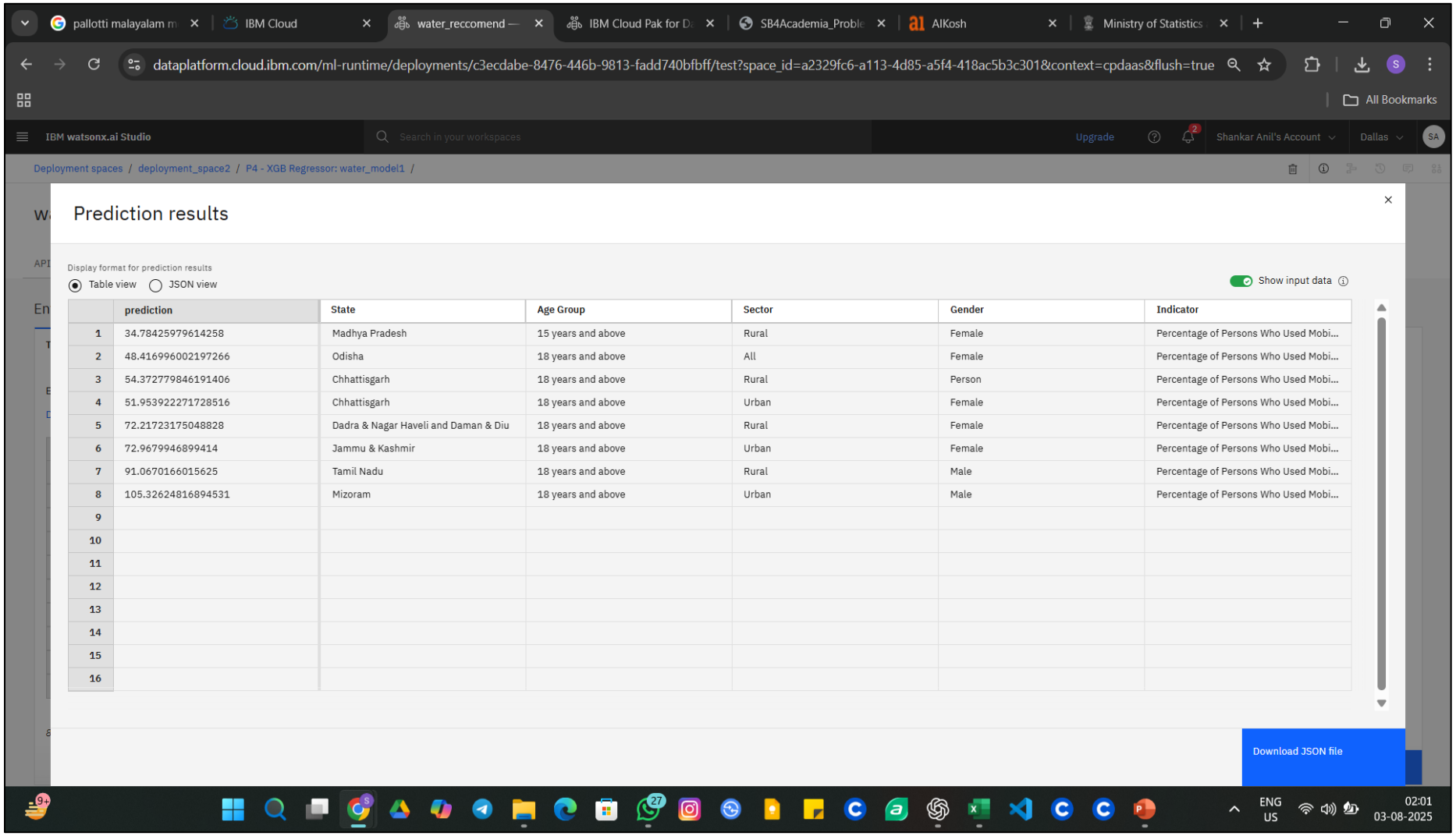
RESULT (SCREENSHOTS)



	State (other)	Age Group (other)	Sector (other)	Gender (other)	Indicator (other)
1	Madhya Pradesh	15 years and above	Rural	Female	Percentage of Persons Who Used Mobile Telephone
2	Odisha	18 years and above	All	Female	Percentage of Persons Who Used Mobile Telephone
3	Chhattisgarh	18 years and above	Rural	Person	Percentage of Persons Who Used Mobile Telephone
4	Chhattisgarh	18 years and above	Urban	Female	Percentage of Persons Who Used Mobile Telephone
5	Dadra & Nagar Haveli and Daman & Diu	18 years and above	Rural	Female	Percentage of Persons Who Used Mobile Telephone
6	Jammu & Kashmir	18 years and above	Urban	Female	Percentage of Persons Who Used Mobile Telephone
7	Tamil Nadu	18 years and above	Rural	Male	Percentage of Persons Who Used Mobile Telephone
8	Mizoram	18 years and above	Urban	Male	Percentage of Persons Who Used Mobile Telephone
9					
10					

edunet foundation

RESULT (SCREENSHOTS)



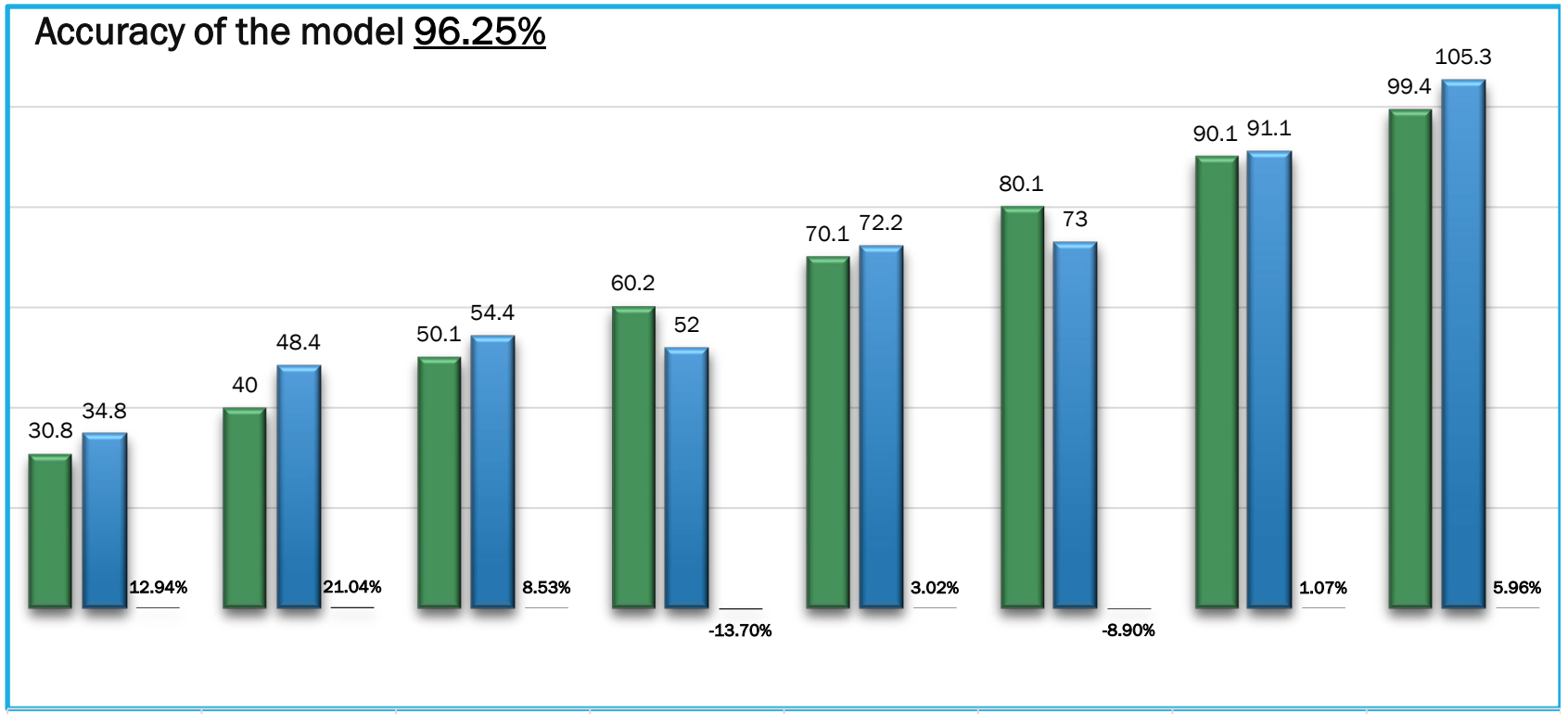
	prediction	State	Age Group	Sector	Gender	Indicator
1	34.78425979614258	Madhya Pradesh	15 years and above	Rural	Female	Percentage of Persons Who Used Mobi...
2	48.416996002197266	Odisha	18 years and above	All	Female	Percentage of Persons Who Used Mobi...
3	54.372779846191406	Chhattisgarh	18 years and above	Rural	Person	Percentage of Persons Who Used Mobi...
4	51.953922271728516	Chhattisgarh	18 years and above	Urban	Female	Percentage of Persons Who Used Mobi...
5	72.21723175048828	Dadra & Nagar Haveli and Daman & Diu	18 years and above	Rural	Female	Percentage of Persons Who Used Mobi...
6	72.9679946899414	Jammu & Kashmir	18 years and above	Urban	Female	Percentage of Persons Who Used Mobi...
7	91.0670166015625	Tamil Nadu	18 years and above	Rural	Male	Percentage of Persons Who Used Mobi...
8	105.32624816894531	Mizoram	18 years and above	Urban	Male	Percentage of Persons Who Used Mobi...
9						
10						
11						
12						
13						
14						
15						
16						

edunet foundation

RESULT (DATA COMPARISON)

State	Age Group	Sector	Gender	Indicator	Value (Dataset)	Predicted Value (in Model)	Change in value (in Percent)
Madhya Pradesh	15 years and above	Rural	Female	Percentage of Persons Who Used Mobile Telephone with Active Sim Card at Least Once During Last Three Months Preceding the Date of the Survey	30.8	34.8	12.94%
Odisha	18 years and above	All	Female	Percentage of Persons Who Used Mobile Telephone with Active Sim Card at Least Once During Last Three Months Preceding the Date of the Survey	40	48.4	21.04%
Chhattisgarh	18 years and above	Rural	Person	Percentage of Persons Who Used Mobile Telephone with Active Sim Card at Least Once During Last Three Months Preceding the Date of the Survey	50.1	54.4	8.53%
Chhattisgarh	18 years and above	Urban	Female	Percentage of Persons Who Used Mobile Telephone with Active Sim Card at Least Once During Last Three Months Preceding the Date of the Survey	60.2	52.0	-13.70%
Dadra & Nagar Haveli and Daman & Diu	18 years and above	Rural	Female	Percentage of Persons Who Used Mobile Telephone with Active Sim Card at Least Once During Last Three Months Preceding the Date of the Survey	70.1	72.2	3.02%
Jammu & Kashmir	18 years and above	Urban	Female	Percentage of Persons Who Used Mobile Telephone with Active Sim Card at Least Once During Last Three Months Preceding the Date of the Survey	80.1	73.0	-8.90%
Tamil Nadu	18 years and above	Rural	Male	Percentage of Persons Who Used Mobile Telephone with Active Sim Card at Least Once During Last Three Months Preceding the Date of the Survey	90.1	91.1	1.07%
Mizoram	18 years and above	Urban	Male	Percentage of Persons Who Used Mobile Telephone with Active Sim Card at Least Once During Last Three Months Preceding the Date of the Survey	99.4	105.3	5.96%
						Avg. Change	3.75%

RESULT (CHART)



	Madhya Pradesh	Odisha	Chhattisgarh	Chhattisgarh	Dadra & Nagar Haveli	Jammu & Kashmir	Tamil Nadu	Mizoram
Value in Dataset	30.8	40	50.1	60.2	70.1	80.1	90.1	99.4
Predicted Value	34.8	48.4	54.4	52	72.2	73	91.1	105.3
Change	12.94%	21.04%	8.53%	-13.70%	3.02%	-8.90%	1.07%	5.96%

CONCLUSION

- The machine learning model developed using demographic and regional features such as State, Age Group, Sector (Urban/Rural), Gender, and Indicator has demonstrated its potential in identifying patterns of inequality in access to essential services, particularly safe drinking water. The inclusion of these socio-demographic attributes has enabled more granular, group-specific insights. For instance, the model effectively highlighted disparities in water access between urban and rural areas, across different age groups and genders, as well as variations among states.
- This approach allows stakeholders to recognize which specific populations are most vulnerable, helping to target interventions more precisely. The model's predictive capacity reinforces the importance of integrating diverse social parameters to understand and address the gaps in resource accessibility, especially in alignment with India’s Sustainable Development Goals (SDGs).

FUTURE SCOPE

- **Integration with Real-Time Data:**
  - Incorporate satellite imagery, IoT-based water sensors, and government APIs for real-time monitoring of water access and quality.
- **Model Expansion:**
  - Extend the model to include other SDG-linked indicators such as sanitation, education, and healthcare for holistic planning.
- **Scalability:**
  - Adapt the model for other developing countries facing similar water accessibility issues.

REFERENCES

- **Dataset Source:**
  - **Improved Source of Drinking Water – Multiple Indicator Survey (78th Round)**
  - **Available at: AI Kosh Dataset Link**
- **Official Statistical Portal:**
  - **e-Sankhyiki – Ministry of Statistics and Programme Implementation (MoSPI)**
  - **Available at: Macro Indicators Portal**

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild
IBM

Shankar Anil

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/746f2002-2870-4324-bc9b-c0e0f6fb041d

IBM

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM

Shankar Anil

Has successfully satisfied the requirements for:

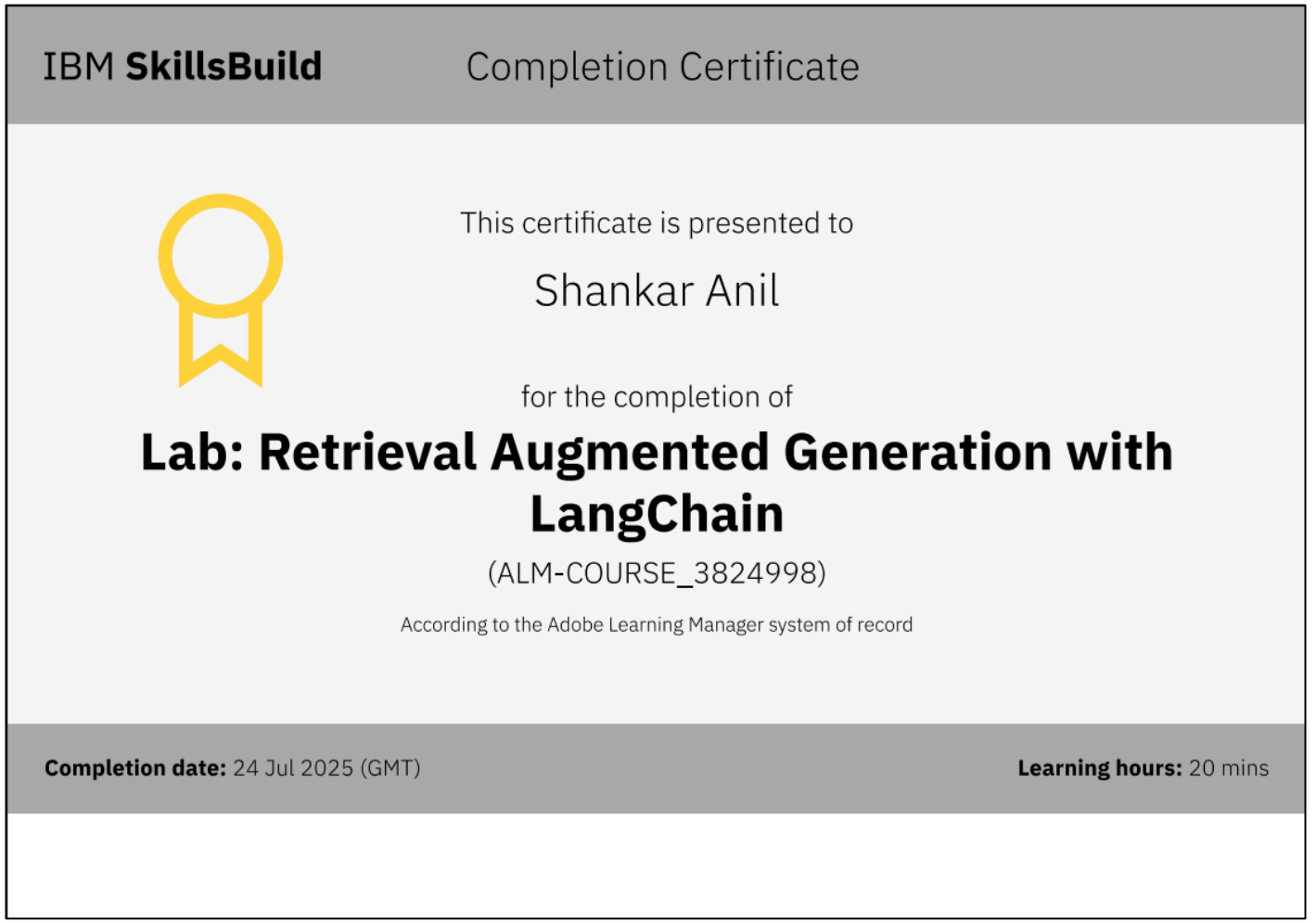
Journey to Cloud: Envisioning Your Solution

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

Verify: https://www.credly.com/badges/9af9a8d6-0afd-49c2-a7d4-65cc5fe4a98b

IBM

IBM CERTIFICATIONS



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Shankar Anil

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

edunet
foundation