CAPSTONE PROJECT

IMPROVED SOURCE OF DRINKING WATER

Presented By:
Vishal.K.R – Sri Venkateswara College of Engineering – CSE

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 NO. 38

Access to safe and improved sources of drinking water remains a critical issue in India, especially in rural and underdeveloped regions. Despite ongoing efforts under the Sustainable Development Goals (SDGs), inequalities persist in water accessibility across states and socio-economic groups. 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. It will also explore related indicators such as use of clean cooking fuel and migration trends. By identifying patterns and disparities, the study will generate actionable insights to support evidence-based policymaking. The ultimate goal is to help ensure equitable access to clean water and contribute to India's progress on SDG targets.

PROPOSED SOLUTION

- Design an **AI-powered analytics platform** to comprehensively analyse drinking water accessibility patterns across India using the 78th Round Multiple Indicator Survey (MIS) data, enabling **data-driven policy recommendations** for equitable water distribution.

Data Collection & Integration:

- Utilize the **78th Round MIS dataset** from AI Kosh containing comprehensive drinking water access indicators.
- Integrate **supplementary datasets** including clean cooking fuel usage and migration patterns.

Data Preprocessing & Feature Engineering:

- **Clean and standardize** multi-source datasets to handle inconsistencies and missing values. Create **composite indicators** combining water access, sanitation, and socio-economic factors.
- **Geographic clustering** to identify regional patterns and disparities. **Temporal analysis** features to track progress over time.

Machine Learning & Analytics:

- Implement **predictive models** (Random Forest, Gradient Boosting) to forecast water accessibility trends
- **Classification algorithms** to categorize regions based on water access levels (High, Medium, Low). **Clustering techniques** (K-means, DBSCAN) to identify similar socio-economic groups
- **Correlation analysis** between water access and related indicators (cooking fuel, migration)

IBM Cloud Deployment Architecture:

- **IBM Watson Studio** for model development and training. **IBM Cloud Object Storage** for secure dataset management and artifacts
- **IBM Cloud Functions** for serverless real-time analytics. **IBM Cognos Analytics** for interactive dashboard creation

Evaluation:

- **Comprehensive Analytics Platform** with real-time monitoring capabilities and **Policy Recommendation Engine** based on data-driven insights
- **Interactive Visualization Dashboard** for stakeholders and **Automated Reporting System** for SDG progress tracking

edunet foundation

SYSTEM APPROACH

System Requirements

- **IBM Cloud Lite Account**→ Services used: Watson Studio, Cloud Object Storage (COS), Cloud Functions
- **Notebook Runtime:** Python 3.10+ (in IBM Watson Studio)
- **Internet Access** for cloud integration and dataset handling

Libraries Used

- **Pandas** → Tabular data processing
- **NumPy** → Efficient numeric computations
- **Matplotlib / Seaborn** → Visual trend analysis
- **ibm-watsonx-ai** → AI integration with foundation models
- **ibm_boto3** → Access IBM Cloud Object Storage programmatically

ALGORITHM & DEPLOYMENT

Algorithm Selection:

- **GroupBy Aggregation** – for computing population proportions. **Chi-Square Analysis** – to test associations (e.g., water source vs. migration)
- **K-Means Clustering** *(Optional)* – to group regions by access similarity

Data Input:

Fields from the **MIS dataset**, including:

- **Primary source of drinking water** and **Type of cooking fuel**
- **Migration details** (household movement patterns). **Demographics:** State, region, income class

Training Process (for clustering):

Data is **preprocessed** (missing values handled, labels encoded)

- Normalize values using **Min-Max scaling**
- Apply **K-Means** with optimal number of clusters chosen via **Elbow Method** and Evaluate cluster quality using **Silhouette Score**

Prediction Process:

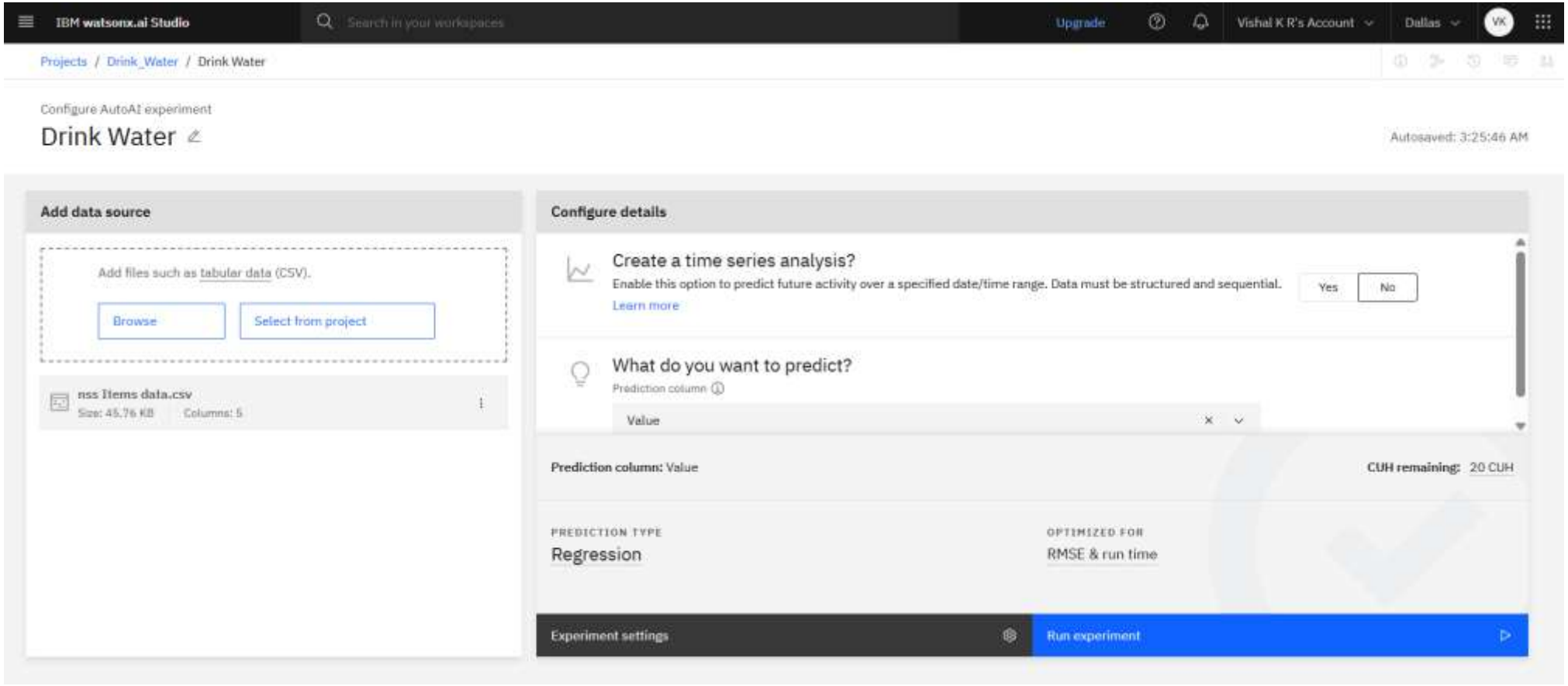
- **Insights** are generated from **visual trends**, **statistical tests**, and **cluster mappings**
- Results are converted into graphs, maps, and policy-friendly summaries
- IBM Watson Studio and Watsonx APIs may assist in summarizing key findings

Deployment:

- All notebooks and analytics were hosted on **IBM Watson Studio (Lite Tier)**
- Final results are made accessible through a **Flask web interface** or triggered via **IBM Cloud Functions**
- Stored and backed up in **IBM Cloud Object Storage (COS)** for reliable access

RESULT

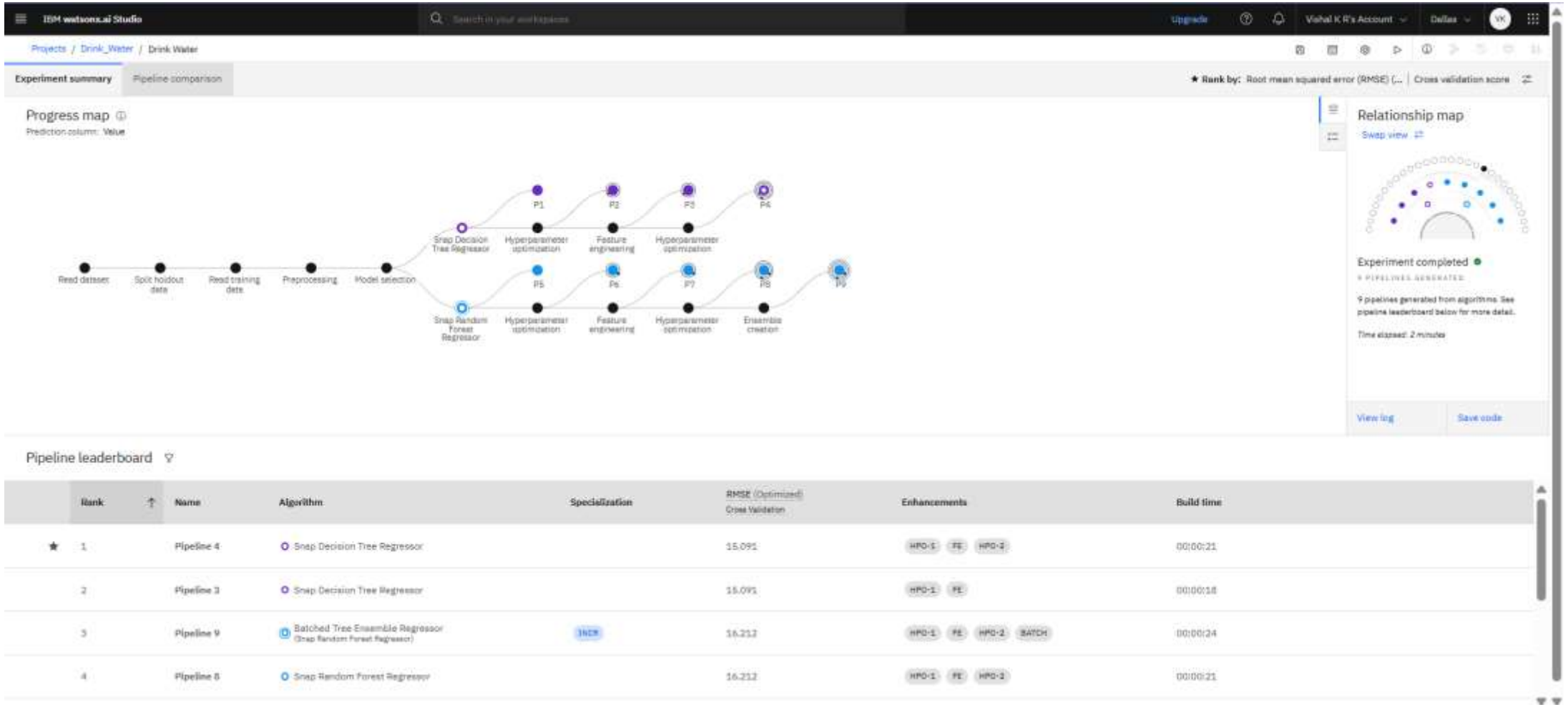
- Creating AutoAI agent.
- Uploading the dataset “nss Items data.csv”.
- Using Prediction Model “Regression” to predict “Value”.



edunet foundation

RESULT

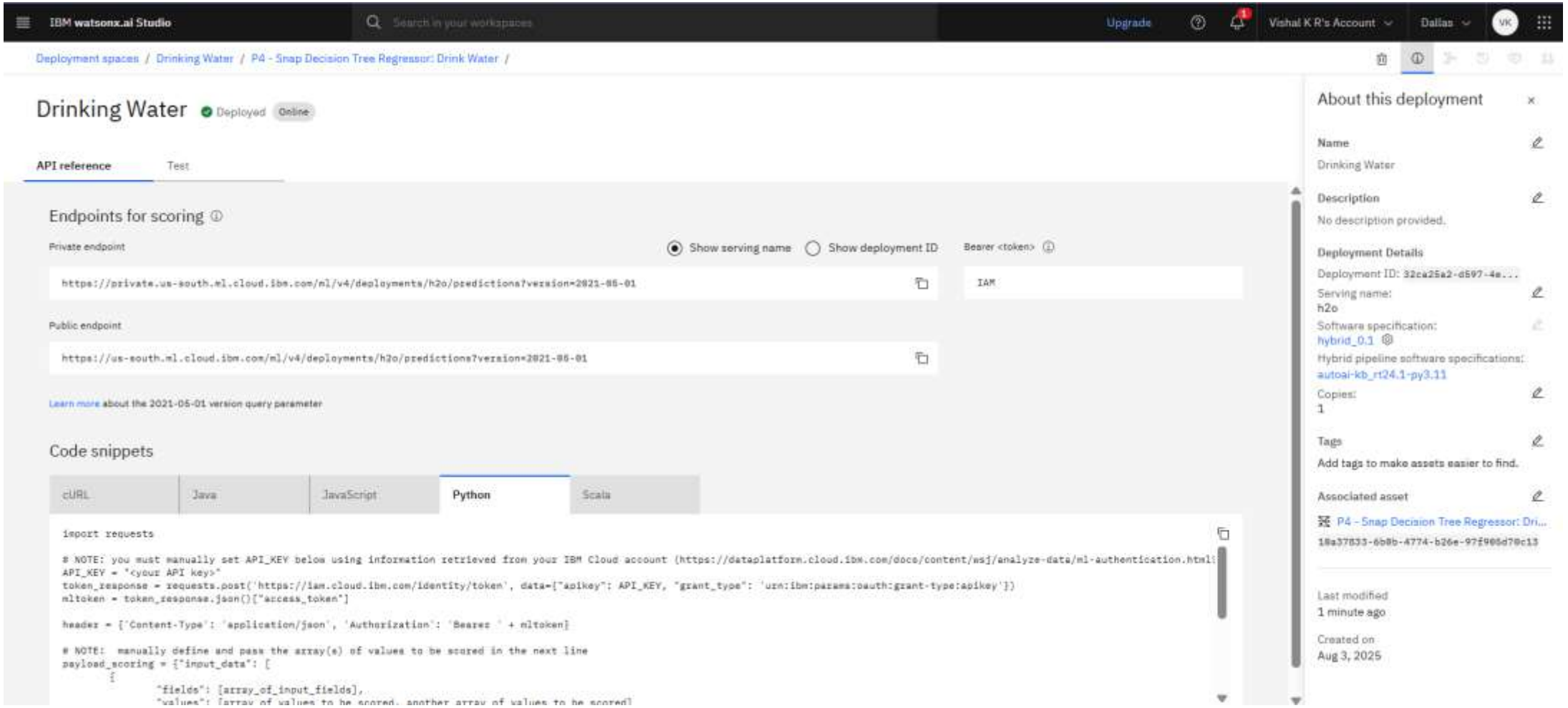
- Training the model and Selecting the best Pipeline “Pipeline 4” as output model.



edunet foundation

RESULT

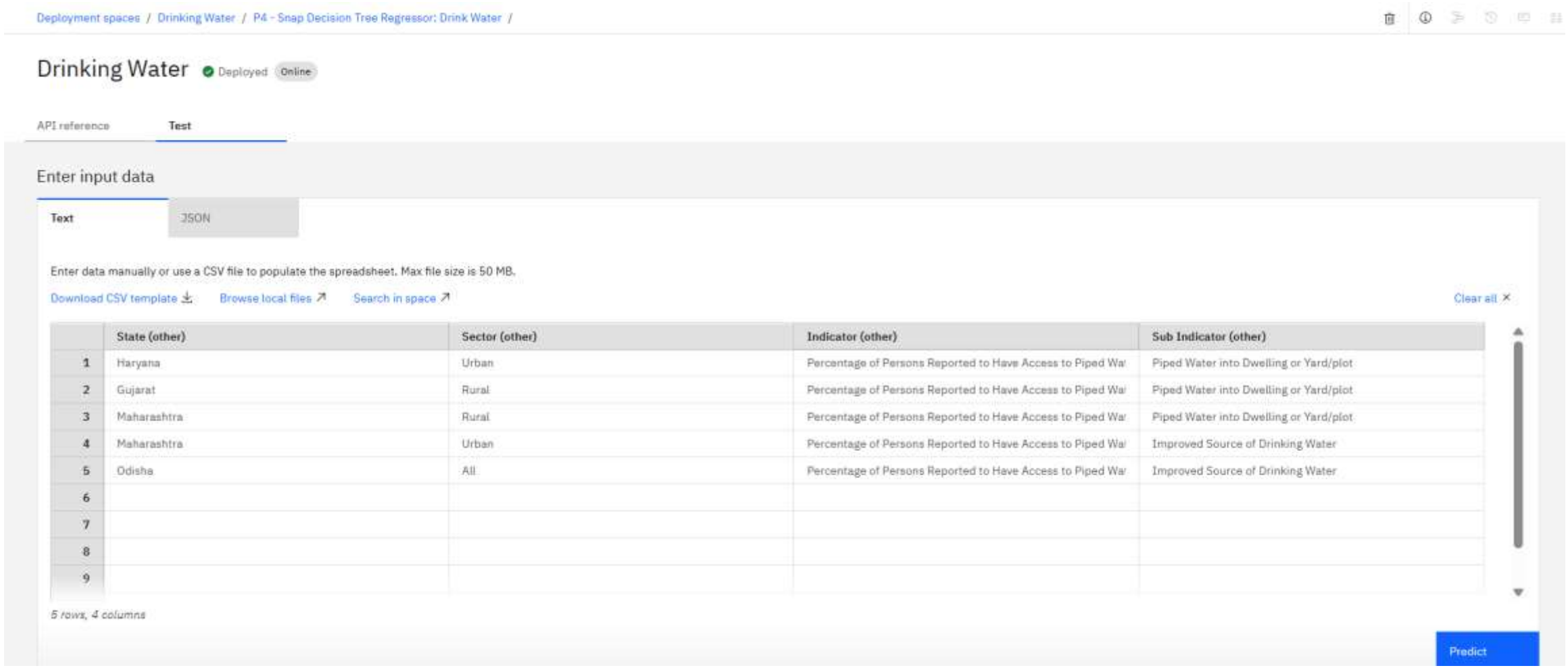
- Deploying the model as “Drinking Water”.



edunet foundation

RESULT

- Loading few Testing Data for prediction.

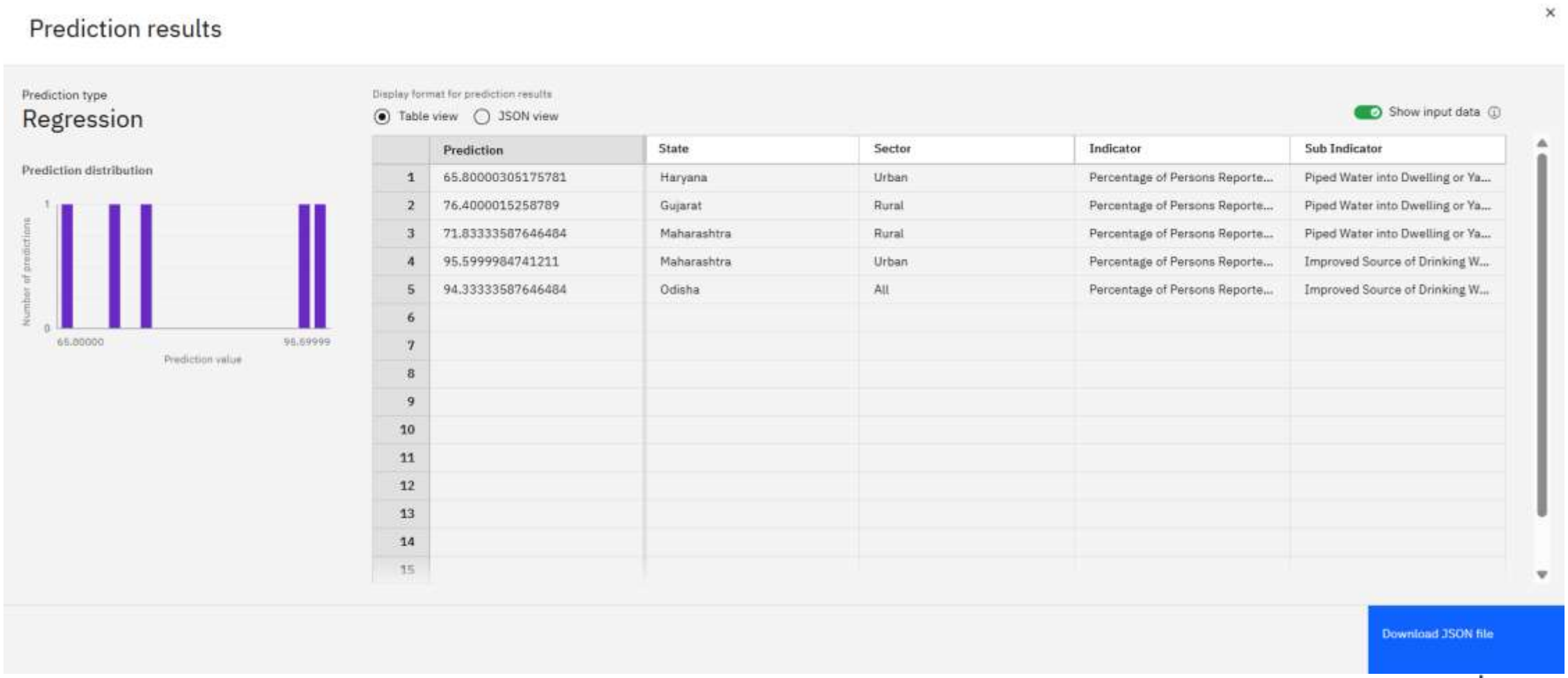


	State (other)	Sector (other)	Indicator (other)	Sub Indicator (other)
1	Haryana	Urban	Percentage of Persons Reported to Have Access to Piped Wa	Piped Water into Dwelling or Yard/plot
2	Gujarat	Rural	Percentage of Persons Reported to Have Access to Piped Wa	Piped Water into Dwelling or Yard/plot
3	Maharashtra	Rural	Percentage of Persons Reported to Have Access to Piped Wa	Piped Water into Dwelling or Yard/plot
4	Maharashtra	Urban	Percentage of Persons Reported to Have Access to Piped Wa	Improved Source of Drinking Water
5	Odisha	All	Percentage of Persons Reported to Have Access to Piped Wa	Improved Source of Drinking Water
6				
7				
8				
9				

edunet foundation

RESULT

- Results of Prediction value with Prediction Distribution.

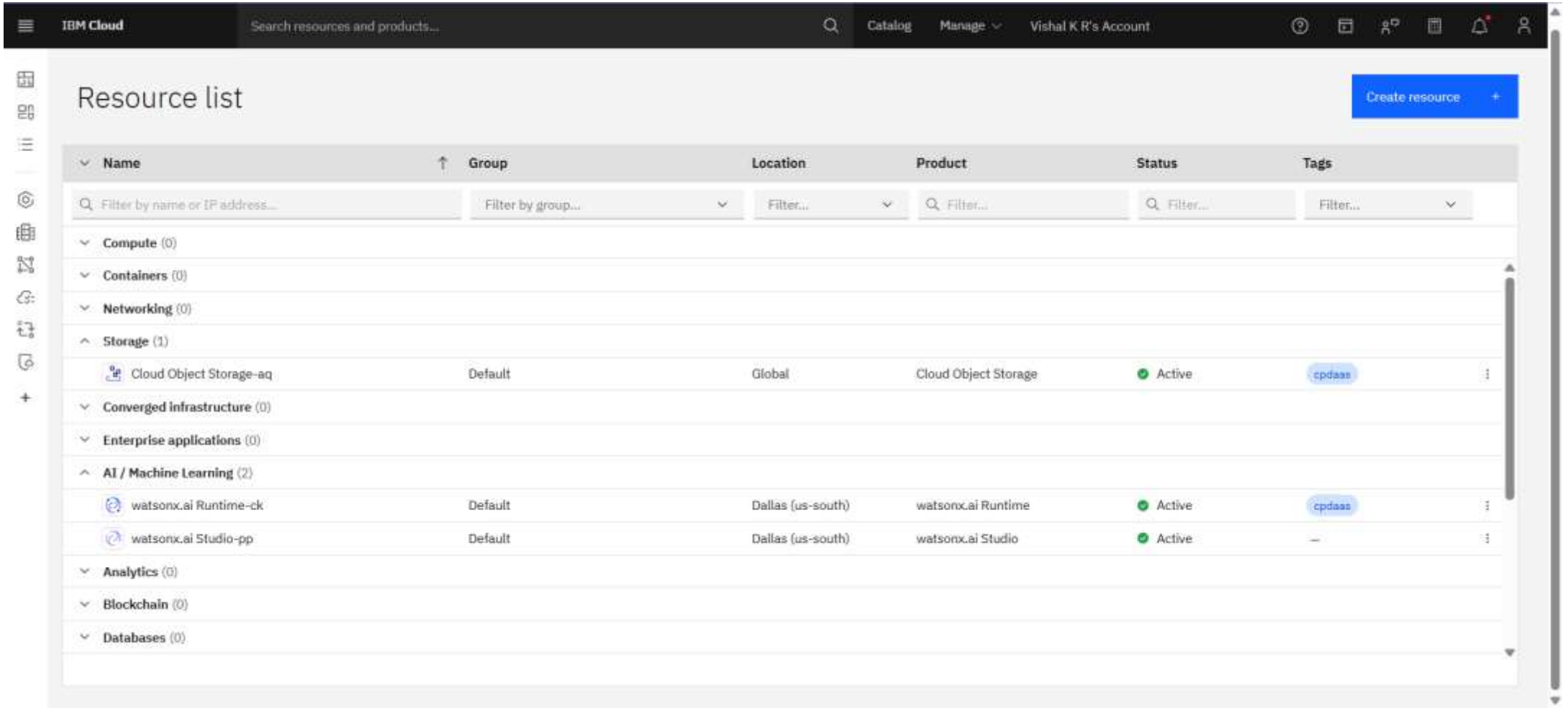


	Prediction	State	Sector	Indicator	Sub Indicator
1	65.80000305175781	Haryana	Urban	Percentage of Persons Reporte...	Piped Water into Dwelling or Ya...
2	76.4000015258789	Gujarat	Rural	Percentage of Persons Reporte...	Piped Water into Dwelling or Ya...
3	71.83333587646484	Maharashtra	Rural	Percentage of Persons Reporte...	Piped Water into Dwelling or Ya...
4	95.5999984741211	Maharashtra	Urban	Percentage of Persons Reporte...	Improved Source of Drinking W...
5	94.33333587646484	Odisha	All	Percentage of Persons Reporte...	Improved Source of Drinking W...
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					

edunet foundation

RESULT

- Resource List



CONCLUSION

Key Findings & Effectiveness:

- The project effectively identified **inequalities in access to improved drinking water** using **government-backed MIS data.**
- Analysis revealed:
  - **State-wise disparities** in improved water access
  - Strong links between **clean fuel access** and **migration patterns**
  - Vulnerable groups (e.g., rural, low-income) with significantly lower access rates

Challenges Faced:

- **Missing and inconsistent entries** in the MIS dataset required intensive cleaning
- Difficulty in **categorical encoding** across multiple social indicators
- Limited availability of **ground-truth labels** for validation

Implementation Success:

- IBM Cloud services like **Watson Studio** and **Cloud Object Storage** enabled seamless processing and storage
- Visualizations and insights were generated efficiently using Python and open-source tools

Impact:

- The project provides a **data-driven lens** for policymakers to target areas with **low water security**
- Supports India's progress toward **Sustainable Development Goal 6 (Clean Water and Sanitation)**

FUTURE SCOPE

- **Incorporate Additional Data Sources**
  Integrate **real-time environmental data**, **socioeconomic indicators**, and **census insights** to enrich the analysis and offer more context-aware recommendations.
- **Optimize Data Processing Pipeline**
  Enhance performance through **automated preprocessing**, **data validation scripts**, and use of **Spark or Pandas on IBM Cloud Pak for Data** to scale efficiently.
- **Expand Regional Coverage**
  Scale the system to analyze **multiple states or union territories**, enabling cross-state comparison and centralized water access planning across India.
- **Advanced Algorithm Enhancement**
  Introduce **advanced ML models** (e.g., ensemble classifiers or transformer-based analytics) for predicting vulnerable regions and policy outcome simulations.
- **Integration with Emerging Technologies**
  Utilize **Edge Computing** (e.g., IBM Edge Application Manager) for on-site deployment in rural areas and explore **Watsonx.ai multimodal models** for richer insights using text + image data.

REFERENCES

- **AI Kosh Dataset**
  Improved Source of Drinking Water – Multiple Indicator Survey (78th Round)
  https://aikosh.indiaai.gov.in/web/datasets/details/improved_source_of_drinking_water_multiple_indicator_survey_78th_round.html
- **IBM Cloud Services**
  IBM Watson Studio, Cloud Object Storage, and IBM Cloud Functions
  https://cloud.ibm.com
- **Watsonx.ai Documentation**
  Official SDK and API reference used for foundation model integration
  https://ibm.github.io/watsonx-ai
- **Sustainable Development Goals (SDG 6 – Clean Water and Sanitation)**
  UN Department of Economic and Social Affairs
  https://sdgs.un.org/goals/goal6
- **Python Libraries**
  Pandas, *Matplotlib*, *Seaborn*, *NumPy*, *Scikit-learn* — used for data analysis and modeling
  https://pypi.org
- **IBM Developer Blog & Tutorials** (for reference on cloud deployment and visual dashboards)
  https://developer.ibm.com

edunet foundation

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

VISHAL K R

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 24, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/fb3c8d2c-c1cd-470f-bca6-de5ea5ab49e9

IBM

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

VISHAL K R CSE

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

edunet foundation

THANK YOU

edunet foundation