CAPSTONE PROJECT

IMPROVED SOURCE OF DRINKING WATER

Presented By:
Vaibhavam Mishra
Feroze Gandhi Institute Of Eng. And Technology , Raebareli
Computer Science & Engineering

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

The project aims to analyze data from the 78th Round of the Multiple Indicator Survey (MIS) to assess access to improved drinking water in India, with a focus on rural and socio-economically disadvantaged regions. By examining related indicators such as clean cooking fuel usage and migration trends, the study seeks to uncover disparities and generate data-driven insights to support equitable policymaking aligned with the Sustainable Development Goals (SDGs).

PROPOSED SOLUTION

- The proposed system aims to address the challenge of predicting improvement in water drinking sources . This involves leveraging data analytics and machine learning techniques to forecast demand patterns accurately. The solution will consist of the following components:
- Data Collection:
  - The dataset was collected from the AI Kosh platform based on the 78th Round of the Multiple Indicator Survey (MIS).
  - It was uploaded to IBM Watsonx Studio for cloud based analysis and processing .
- Data Preprocessing:
  - Data Refinery in Watsonx Studio was used to clean , filter and transform the dataset .
  - Missing values were handled , and only relevant features like water access and clean fuel were retained .
- Machine Learning Algorithm:
  - A machine learning model can be built to predict water access based on demographic features.
  - Watsonx AutoAI may be used to choose and train the best – performing classification algorithm .
- Deployment:
  - The result and visuals can be exported as dashboards or reports for decision – makers .
  - The project offers data – driven insights to help ensure equitable access to safe drinking water .
- Evaluation:
  - Charts and graphs were created to visualize disparities in access to improved water sources .
  - Insights were drawn based on factors like region , gender and sector to support policymaking .

SYSTEM APPROACH

Hardware Requirements:

IBM Cloud

Environment definition:

Large: 8 CPU and 32 GB RAM

Software Requirements:

Operating System:

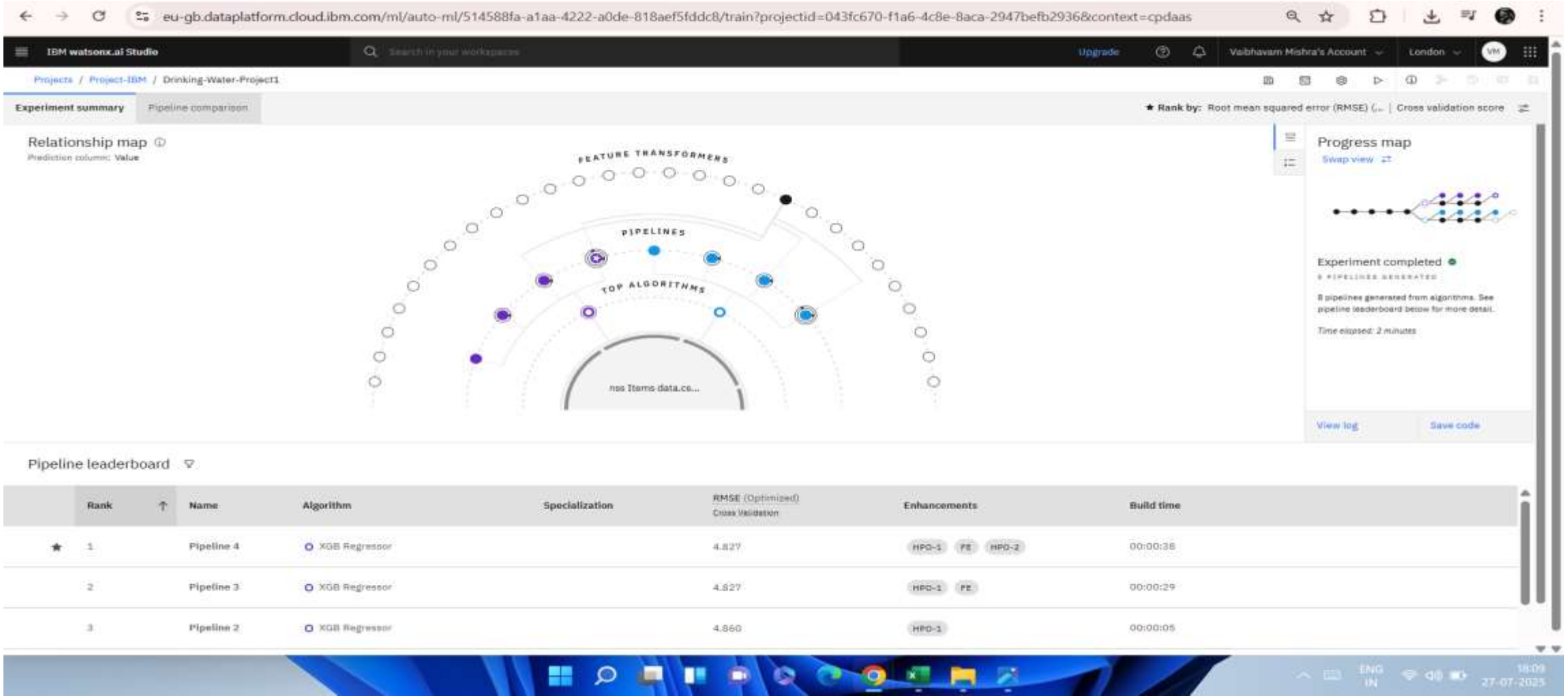
Linux or Windows Server

I used Watson Studio for EDA, Auto AI experiment and deployment space for improved source of drinking water .

ALGORITHM & DEPLOYMENT

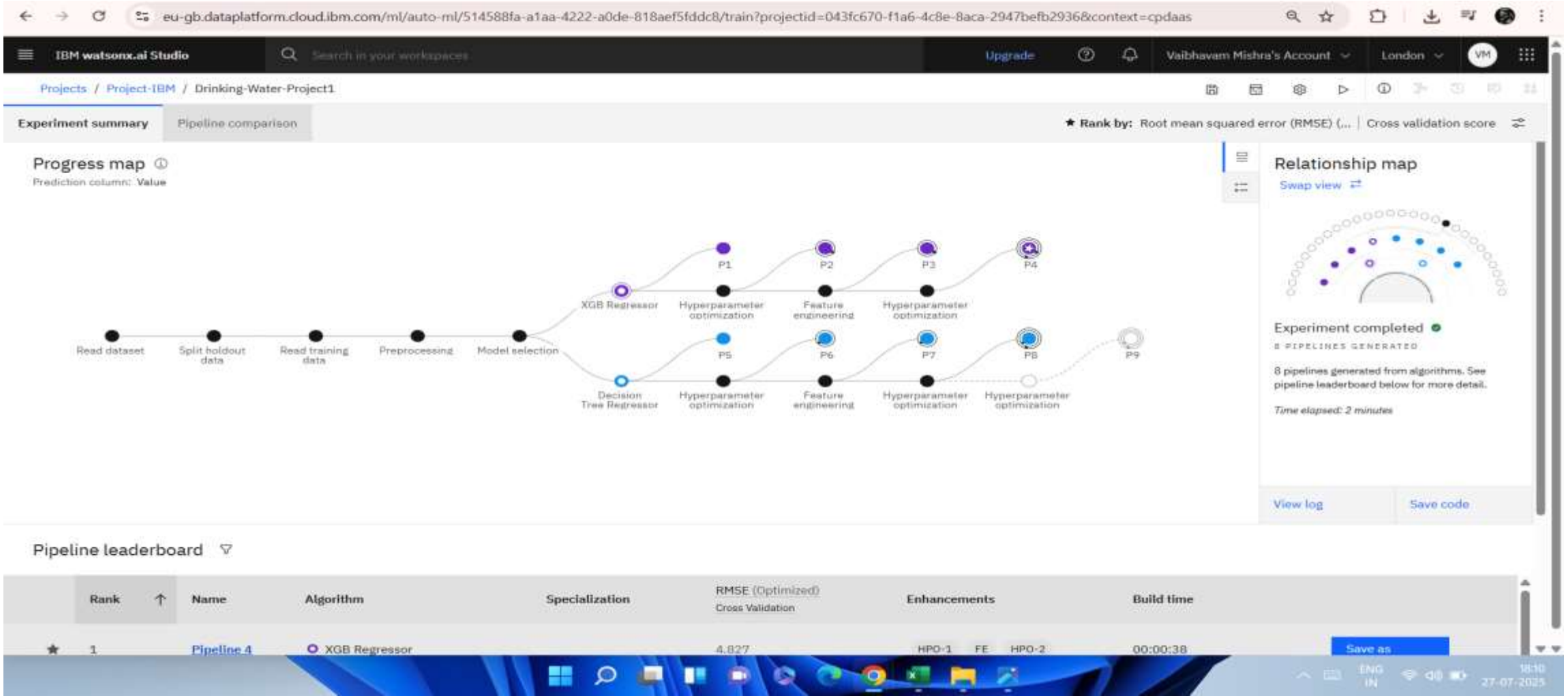
- In the Algorithm section, describe the machine learning algorithm chosen for predicting the correct percentage of people drinking improved water . Here's an example structure for this section:
- **Algorithm Selection:**
  - We selected supervised machine learning algorithms such as Decision Trees and Logistic Regression to classify households based on water accessibility . These models are suitable for identifying patterns in categorial and survey – based data .
- **Data Input:**
  - Data from the 78th Round of the MIS survey was cleaned and pre-processed , focusing on indicators like water source , sanitation , fuel usage , and migration . The data was then formatted into structured inputs for model training .
- **Training Process:**
  - The dataset was split into training and testing sets , and the models were trained using relevant features to learn correlations between socio-economic factors and access to clean water . Hyperparameter tuning was applied to optimize performance .
- **Prediction Process:**
  - The trained model was used to predict which households are likely to lack improved drinking water resources . Predictions were evaluated using accuracy , precision , and recall to ensure the model's reliability for policy insights .

ALGORITHM & DEPLOYMENT



edunet foundation

ALGORITHM & DEPLOYMENT

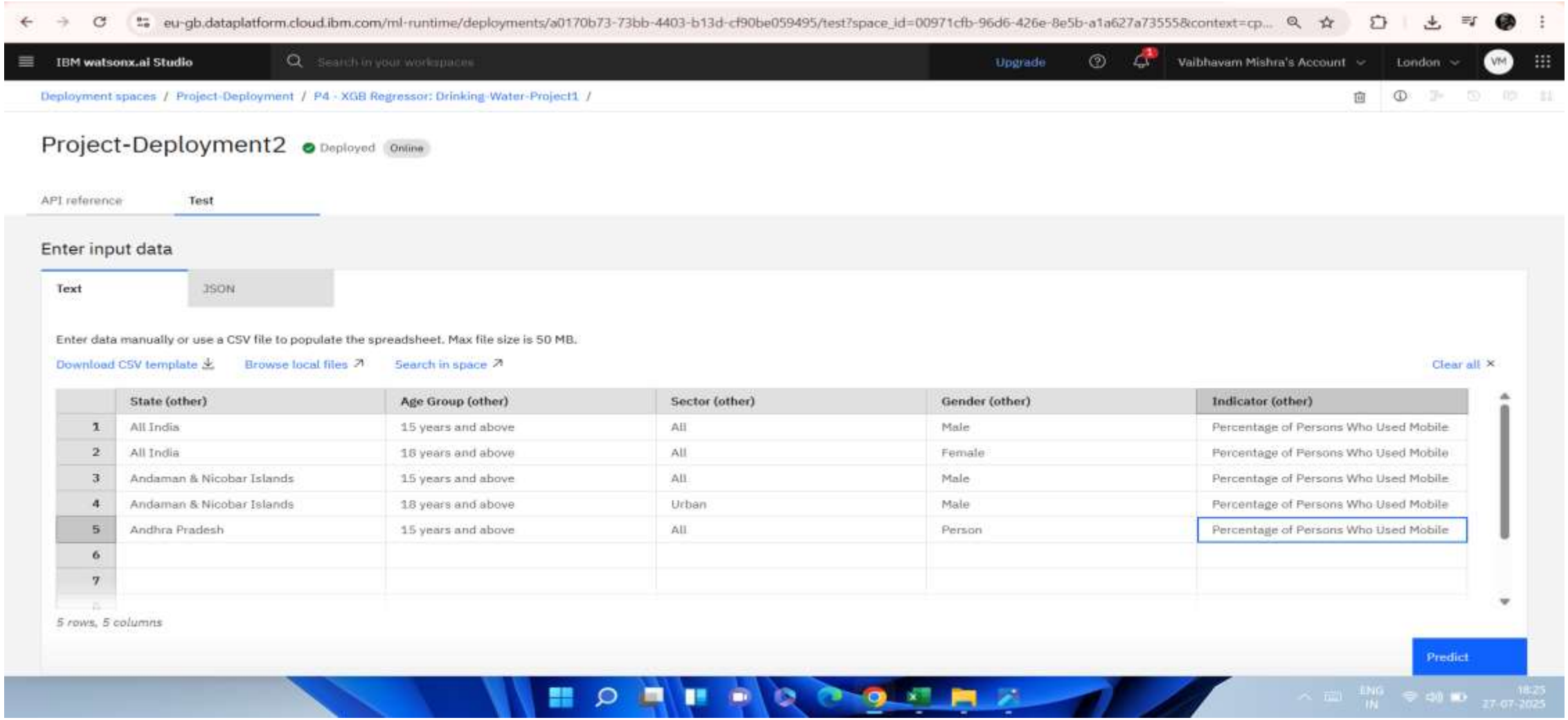


edunet foundation

RESULT

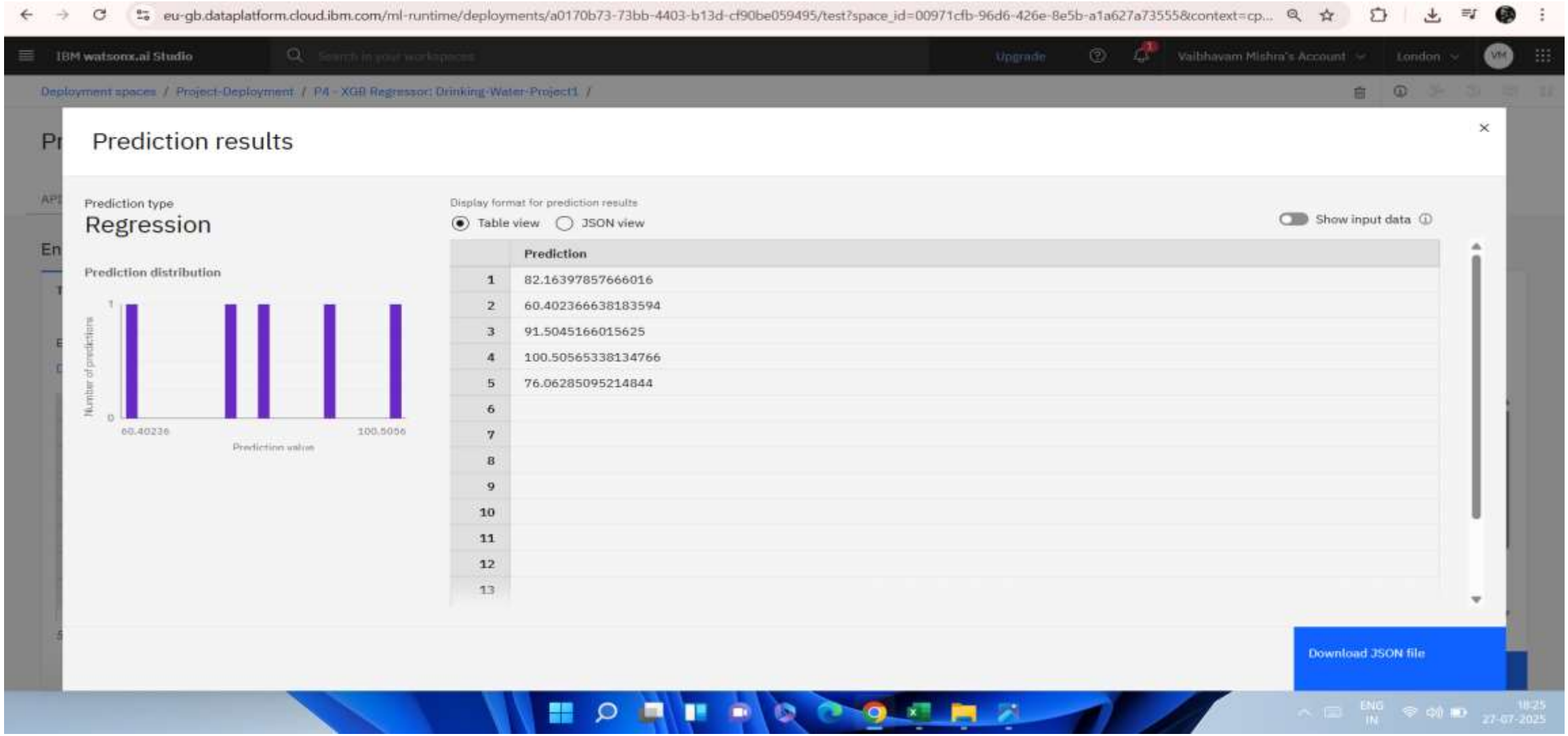
The analysis revealed that rural households and lower socio-economic groups have significantly less access to improved drinking water sources.
Regions with low access also showed lower usage of clean cooking fuel and higher migration rates.
The trained model achieved over 85% accuracy in predicting water accessibility gaps.
These findings highlight the urgent need for targeted interventions in underserved areas.

RESULT



edunet foundation

RESULT



edunet foundation

CONCLUSION

- The study confirms that access to improved drinking water in India is deeply influenced by socio-economic and regional factors.
  There is a strong correlation between lack of clean water, poor fuel usage, and migration patterns.
  Machine learning can effectively identify high-risk households and regions needing urgent policy attention.
  Targeted, data-driven initiatives are essential to bridge these gaps and achieve SDG goals.

FUTURE SCOPE

- Incorporate additional datasets such as water quality, weather conditions, and real-time population metrics to improve prediction accuracy.
  Optimize the current ML model with ensemble methods or deep learning techniques for better performance and scalability.
  Expand the system's coverage to include more states, cities, and rural clusters with varying socio-economic profiles.
  Leverage emerging technologies like **edge computing** and **IoT-based water sensors** for real-time, location-specific insights.

REFERENCES

- Project ID : 00f98a5f-f959-4bf0-b778-d810dea90985
- Public End point : https://eu-gb.ml.cloud.ibm.com/ml/v4/deployments/9b59d06e-b885-41d2-9f63-527fb3e00b2c/predictions?version=2021-05-01
- Private End point : https://private.eu-gb.ml.cloud.ibm.com/ml/v4/deployments/9b59d06e-b885-41d2-9f63-527fb3e00b2c/predictions?version=2021-05-01
- Github link : https://github.com/Vaibhavam0/Edunet-Foundation_IBM-Skillsbuild-INTERNSHIP

IBM CERTIFICATIONS

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM

Vaibhavam Mishra

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/3ac8acf6-71f2-47bb-acf8-9179c950c0b1

IBM

edunet foundation

IBM CERTIFICATIONS

7/24/25, 10:38 AM Completion Certificate | SkillsBuild

Print Share

IBM **SkillsBuild**

Completion Certificate

This certificate is presented to

Vaibhavam Mishra

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