CAPSTONE PROJECT

IMPROVED SOURCE OF DRINKING WATER

Presented By:
1. Gonu Manish Reddy - Dayananda Sagar University - AIML

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

Many people in India, especially in rural and less developed areas, still do not have access to safe and clean drinking water. Even though the government is working towards improving this through the Sustainable Development Goals (SDGs), there are still big differences in water access across states and among different social and economic groups. Other factors like the type of cooking fuel used and people moving from one place to another (migration) may also affect water access. It is important to study this data to understand where the problems are and which groups are most affected.

PROPOSED SOLUTION

- Develop a data-driven system to analyze access to improved drinking water across India using survey data from the 78th Round of the Multiple Indicator Survey (MIS) The system will process household level and regional data to identify patterns and disparities in water access. Insights generated from this analysis will support evidence-based decision making for water infrastructure planning and help meet Sustainable Development Goals (SDGs).

- Data Collection:
  - Use the official dataset from the 78th Round of the Multiple Indicator Survey (MIS) available via Al Kosh.

- Data Preprocessing:
  - Clean the data by removing missing or incorrect values and Rename confusing column names to simple, clear labels.
  - engineer features such as region wise access rates and social economic groupings.

- Machine Learning Algorithm:
  - Implement a machine learning algorithm, Use classification models (e.g., decision trees, logistic regression) to predict likelihood of improved water access based on socio-economic features.

- Evaluation:
  - Check how accurate and useful the model or dashboard is and Use basic metrics like accuracy or visual clarity.
  - Fine-tune the model based on feedback and continuous monitoring of prediction accuracy.

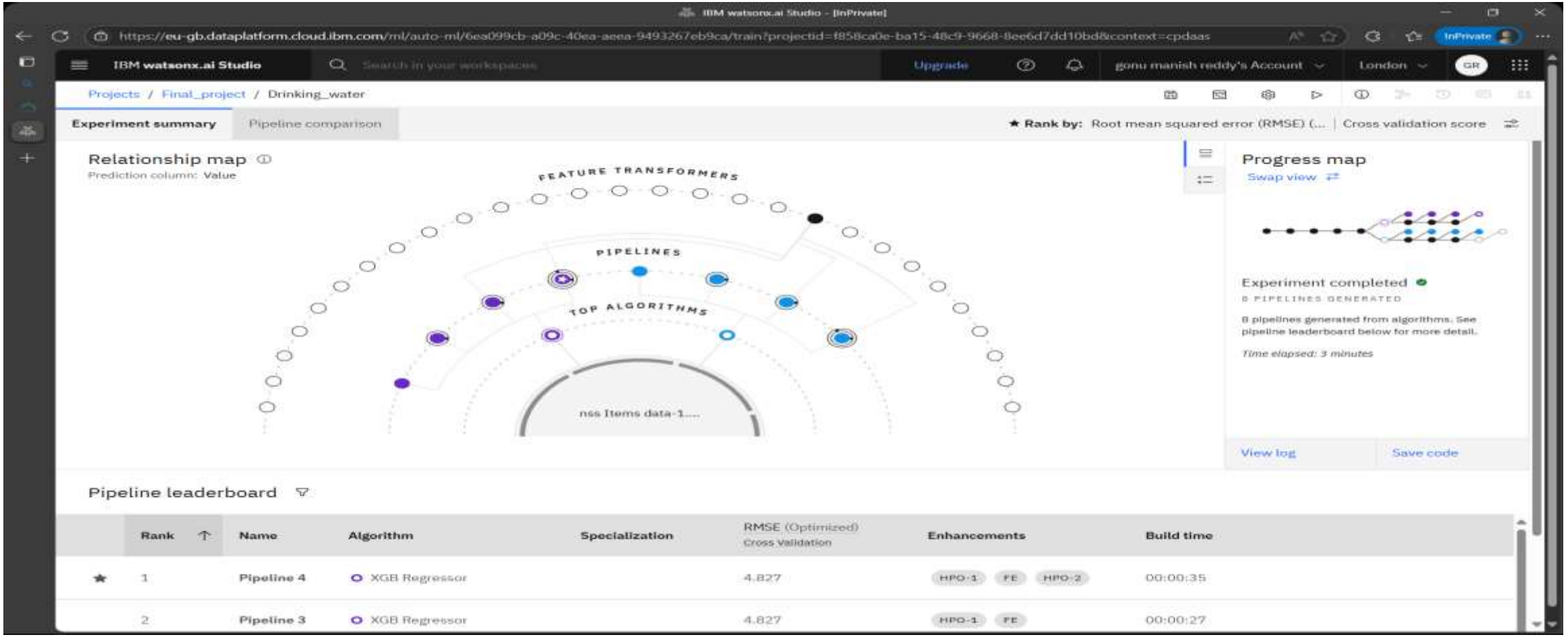
SYSTEM APPROACH

- IBM Cloud(mandatory)
- IBM Watson studio for model development and deployment
- IBM cloud object storage for dataset handling.
- IBM AutoAI for automated ML pipeline.
- ChatGPT for content structuring and research guidance

ALGORITHM & DEPLOYMENT

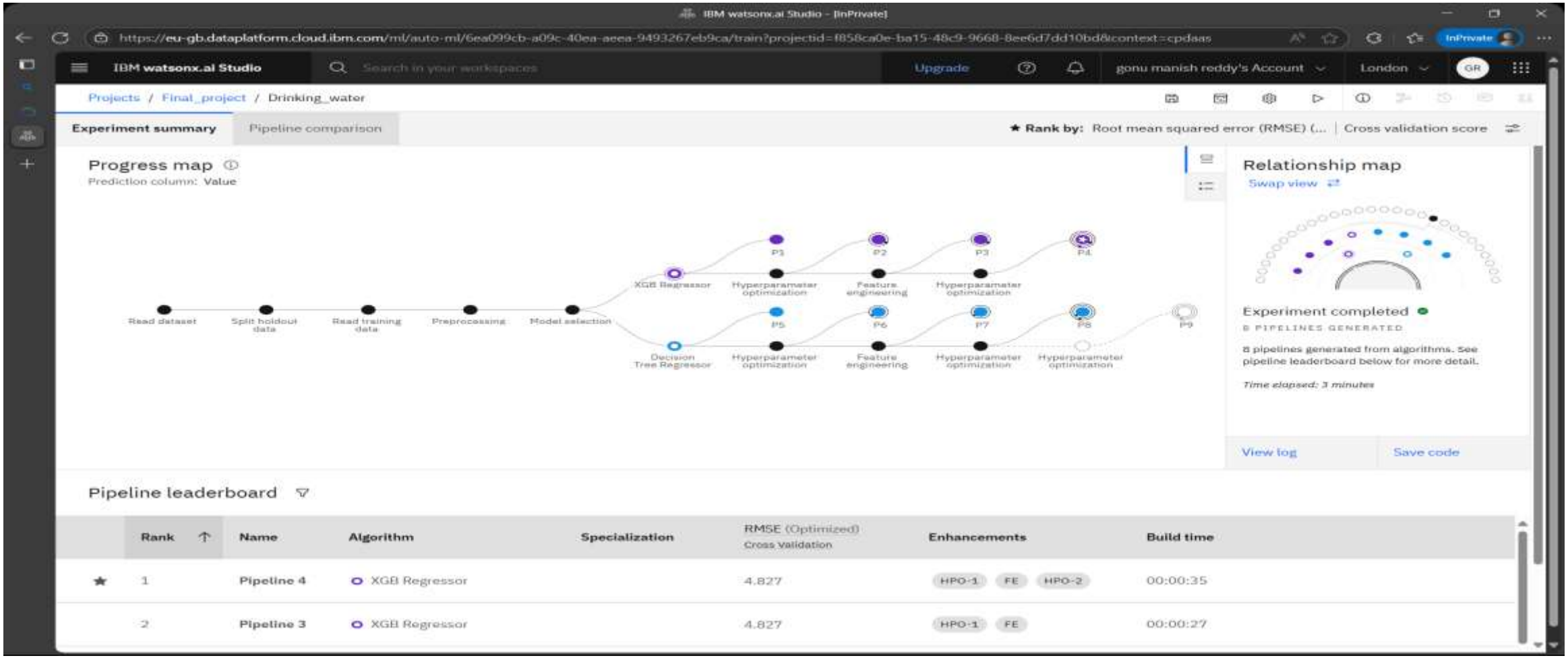
- Algorithm Selection:
  - We chose classification algorithms like XGBoost Classifier to predict whether a household has access to improved drinking water. These models are good for this type problems and work well with survey data.
- Data Input:
  - The model uses inputs like State, Age Group, Sector, Gender and Indicator to make predictions. The target output is whether the household has access to improved water with the percentage.
- Training Process:
  - We split the data into training and testing parts and use techniques like cross-validation to avoid overfitting. Categorical data is converted into numbers so the model can understand it better and the data is trained by auto code in IBM cloud.
- Prediction Process:
  - After training, the model can take new household data and predict if they have access to improved drinking water. This helps in finding areas or groups that need better water access.

RESULT



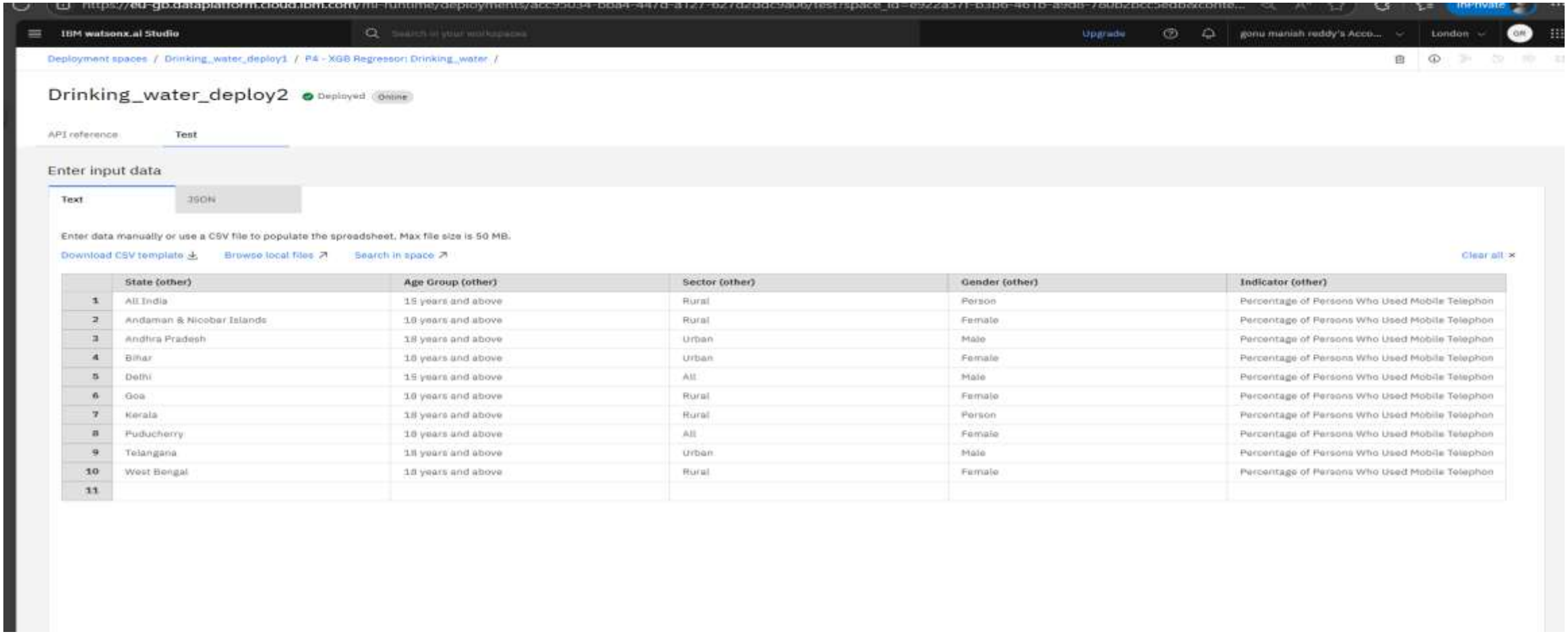
Relationship map

RESULT



Progress map

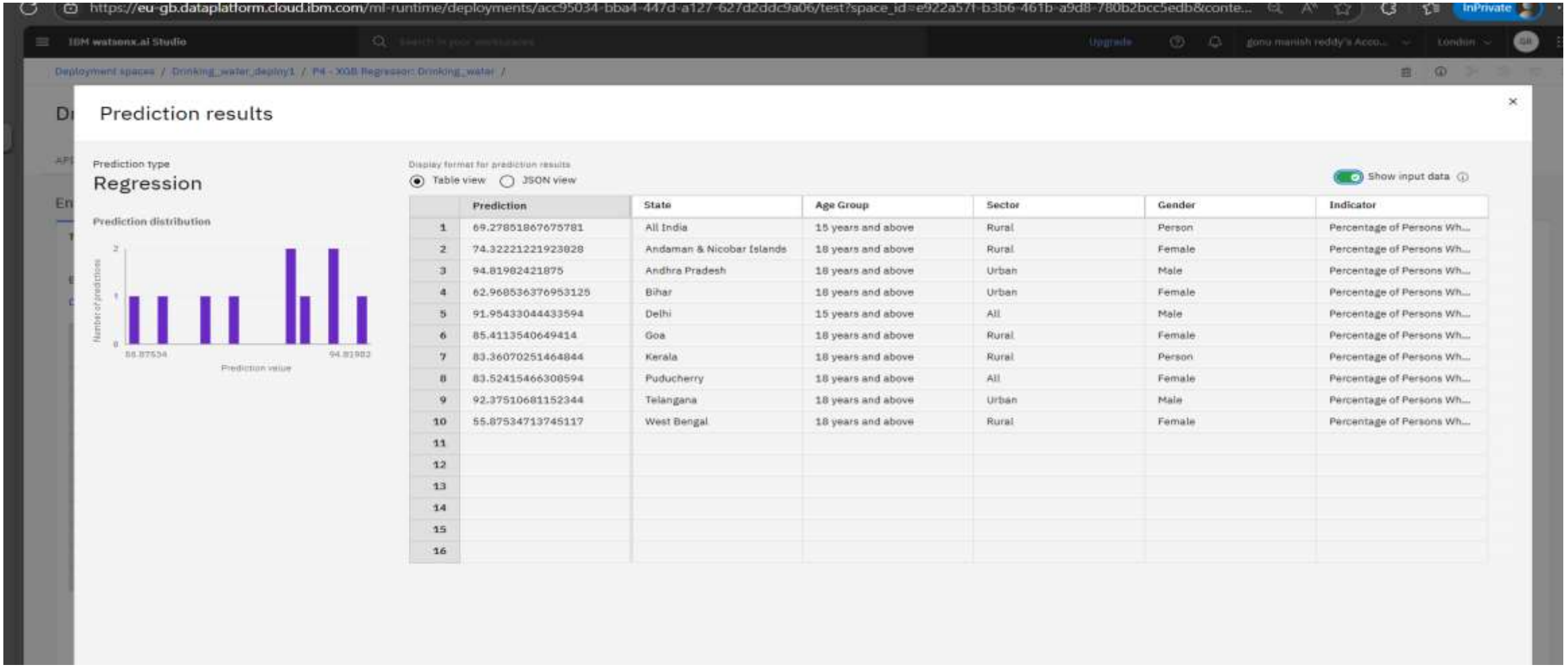
RESULT



	State (other)	Age Group (other)	Sector (other)	Gender (other)	Indicator (other)
1	All India	15 years and above	Rural	Person	Percentage of Persons Who Used Mobile Telephon
2	Andaman & Nicobar Islands	18 years and above	Rural	Female	Percentage of Persons Who Used Mobile Telephon
3	Andhra Pradesh	18 years and above	Urban	Male	Percentage of Persons Who Used Mobile Telephon
4	Bihar	18 years and above	Urban	Female	Percentage of Persons Who Used Mobile Telephon
5	Delhi	15 years and above	All	Male	Percentage of Persons Who Used Mobile Telephon
6	Goa	18 years and above	Rural	Female	Percentage of Persons Who Used Mobile Telephon
7	Kerala	18 years and above	Rural	Person	Percentage of Persons Who Used Mobile Telephon
8	Puducherry	18 years and above	All	Female	Percentage of Persons Who Used Mobile Telephon
9	Telangana	18 years and above	Urban	Male	Percentage of Persons Who Used Mobile Telephon
10	West Bengal	18 years and above	Rural	Female	Percentage of Persons Who Used Mobile Telephon
11					

Input data for predation

RESULT



	Prediction	State	Age Group	Sector	Gender	Indicator
1	69.27851867675781	All India	15 years and above	Rural	Person	Percentage of Persons Wh...
2	74.32221221923828	Andaman & Nicobar Islands	18 years and above	Rural	Female	Percentage of Persons Wh...
3	94.81982421875	Andhra Pradesh	18 years and above	Urban	Male	Percentage of Persons Wh...
4	62.968536376953125	Bihar	18 years and above	Urban	Female	Percentage of Persons Wh...
5	91.95433044433594	Delhi	15 years and above	All	Male	Percentage of Persons Wh...
6	85.4113540649414	Goa	18 years and above	Rural	Female	Percentage of Persons Wh...
7	83.36070251464844	Kerala	18 years and above	Rural	Person	Percentage of Persons Wh...
8	83.52415466308594	Puducherry	18 years and above	All	Female	Percentage of Persons Wh...
9	92.37510681152344	Telangana	18 years and above	Urban	Male	Percentage of Persons Wh...
10	55.87534713745117	West Bengal	18 years and above	Rural	Female	Percentage of Persons Wh...
11						
12						
13						
14						
15						
16						

Prediction Result

edunet foundation

CONCLUSION

We created a system that effectively analyzes survey data to predict who has clean drinking water access. Using IBM Cloud and Watson Studio's automated tools, we could quickly process large amounts of information and build accurate prediction models without needing complex coding. This means we can efficiently turn raw data into valuable insights. Despite some early challenges with data quality, our project demonstrates how cloud based AI can directly support evidence based policy making and predicting clean water access is a crucial step towards achieving sustainable development and improving quality of life in underserved regions, allowing for more effective allocation of resources.

FUTURE SCOPE

In the future, this system can be improved by adding more data from different surveys or government sources to get a better and bigger picture of water access. The prediction model can be made more accurate by trying advanced machine learning techniques. The system can also be expanded to cover more states, cities, or even the whole country. We can also explore new technologies like real-time data collection and edge computing to make the system faster and smarter. These improvements will help make the solution more useful for planning and decision making.

REFERENCES

- The dataset used is from the 78th Round of the Multiple Indicator Survey (MIS), provided by the Ministry of Statistics and Programme Implementation (MoSPI), Government of India.
- The data was accessed via the AI Kosh platform, a national repository for AI datasets in India.
- IBM Cloud Lite services, including Watson Studio and AutoAI, were used to process the data and build the predictive model.
- Official IBM documentation and tutorials supported the setup, training, and deployment of the model.
- Best practices in data preprocessing, model evaluation, and AI for social good were followed based on publicly available case studies and guides.
- ChatGPT was used as a research assistant to explore ideas, write content, and improve project clarity and presentation.
- This project was completed under the guidance of Narendra Eluri , whose support and mentorship played a key role in shaping the solution.

IBM CERTIFICATIONS

IBM CERTIFICATIONS

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

gonu manish Reddy

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