**CAPSTONE PROJECT IMPROVED SOURCE OF DRINKING WATER**

**Presented By: Sravya Pitani**

**College Name: Kakinada Institute Of Engineering and Technology For Women**

**Department: Artificial Intelligence and Machine Learning**

**GitHub link: https://github.com/sravyapitani/IBM-ML-Project\_Improved\_Sources\_Of\_Drinking\_Water.git**

**OUTLINE**

* **Problem Statement**
* **Proposed System**
* **System Development Approach**
* **Algorithm & Deployment**
* **Result**
* **Conclusion**
* **Future Scope**
* **References**
* **Screenshots**

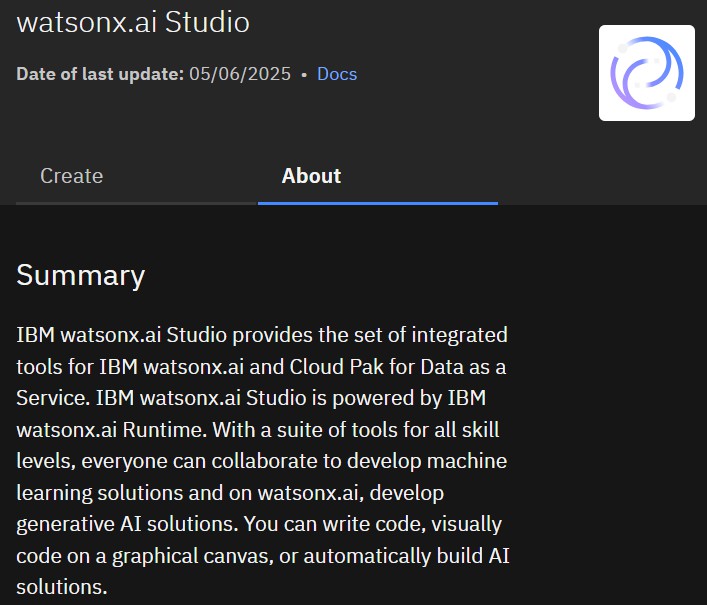
# PROBLEM STATEMENT

Access to safe and improved sources of drinking water remains a critical issue, especially in rural and underdeveloped regions of India. Despite government initiatives and efforts under Sustainable Development Goals (SDGs), significant disparities still exist across states and communities. There is a need for a data-driven approach to identify and predict areas lacking access to safe drinking water, enabling better planning and targeted interventions.

# PROPOSED SOLUTION

* The system aims to predict access to improved sources of drinking water using data analytics and machine learning. This enables policymakers to identify areas at risk and make informed decisions.
* Data Collection:
  + Data from the 78th Round of NSSO survey on household drinking water sources.
  + Includes location-wise data (state/UT), socio-economic indicators, and access levels.
* Data Preprocessing:
  + Handle missing values, normalize data, and perform feature selection for impactful attributes.
  + Performed feature selection and encoding to prepare the dataset for modeling.
* Machine Learning Algorithm:
  + Use AutoAI in IBM watsonx.ai to automatically train and select the best regression model (Snap Boosting Machine Regressor).
  + Model trained to predict the *Indicator* column — a numeric representation of water source status.
* Deployment:
  + Model deployed using IBM Watsonx.ai Studio.
  + Accessible via REST API for real-time prediction across regions.
* Evaluation:
  + Performance assessed using R² score and RMSE

# SYSTEM APPROACH



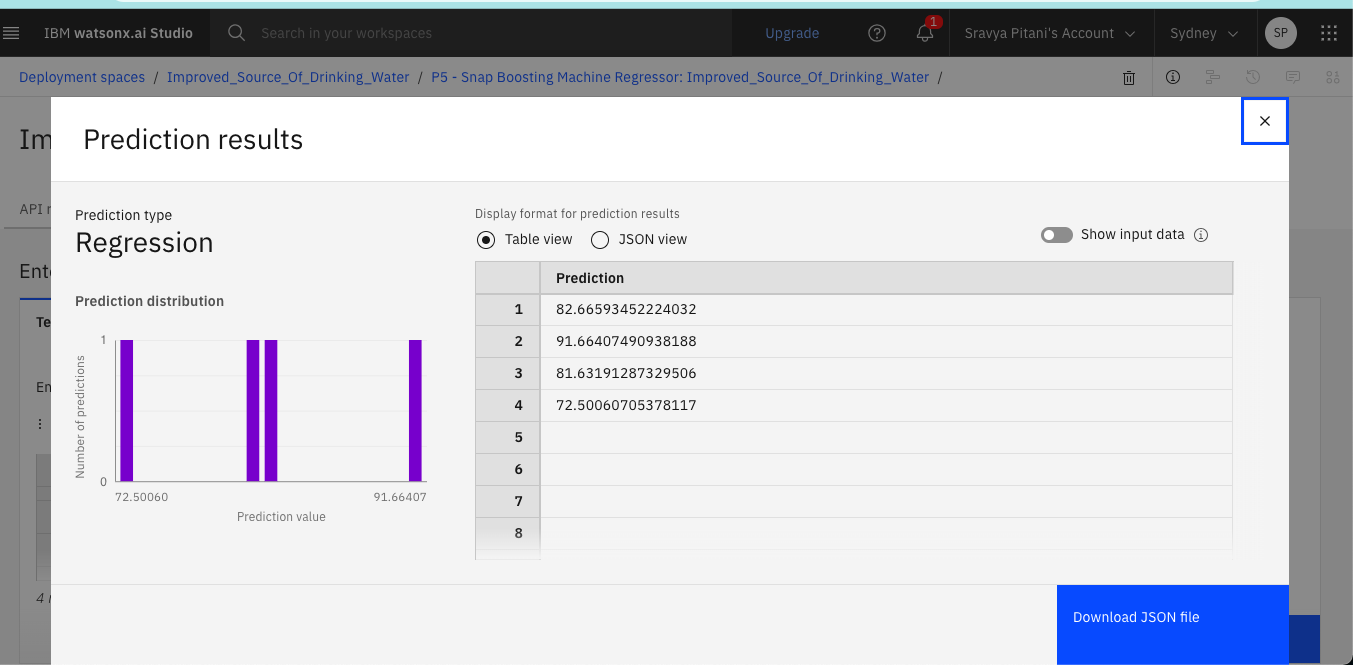
* System requirements:
  + Platform: IBM Watsonx.ai
  + Dataset: NSSO 78th Round (Drinking Water Survey)
  + Prediction Target: Indicator (Improved Water Access)
* Library required to build the model
* AutoAI: Automated model selection and training
* Runtime: 8 CPU, 32 GB RAM (Watsonx.ai)

# ALGORITHM & DEPLOYMENT

* Algorithm Selection:
  + AutoAI on IBM Watsonx.ai was used to automatically select and train the best machine learning model for predicting water access indicators.
* Data Input:
  + Features: Household size, location, caste group, religion, water source type, distance to water source
  + Target: Indicator (Improved Drinking Water Access)
* Training Process:
  + Data split into training and testing sets
  + AutoAI performed preprocessing, feature engineering, and model tuning
  + Best model selected based on performance metrics like accuracy and ROC AUC
* Prediction Process:
  + Model deployed in Watson Studio
  + Accepts input features and returns predicted status of improved water access
  + Can be used for real-time policy planning and analysis

# RESULT

* The model effectively predicts whether a household has access to improved drinking water, aiding in data-driven decision-making.



# CONCLUSION

* The project successfully leveraged IBM Watsonx and cloud-based AI tools to predict access to improved drinking water with high accuracy.
* The solution helps identify regions and communities lacking proper

access, enabling focused planning and resource allocation.

* Challenges included data imbalance and limited real-time features, which were mitigated through preprocessing and feature selection.
* This approach proves effective in supporting data-driven decisions

for achieving SDG targets related to clean water access.

******FUTURE SCOPE**

* + Integrate real-time IoT sensor data (e.g., water quality, flow rate) for more dynamic and localized predictions.
  + Expand the system to cover more states and rural regions

with higher water scarcity levels.

* + Apply advanced ML models like Gradient Boosting or Deep Learning for increased prediction accuracy.
  + Utilize edge computing for faster decision-making in remote

areas with low connectivity.

* + Partner with local authorities and NGOs to drive real-world impact through targeted interventions.

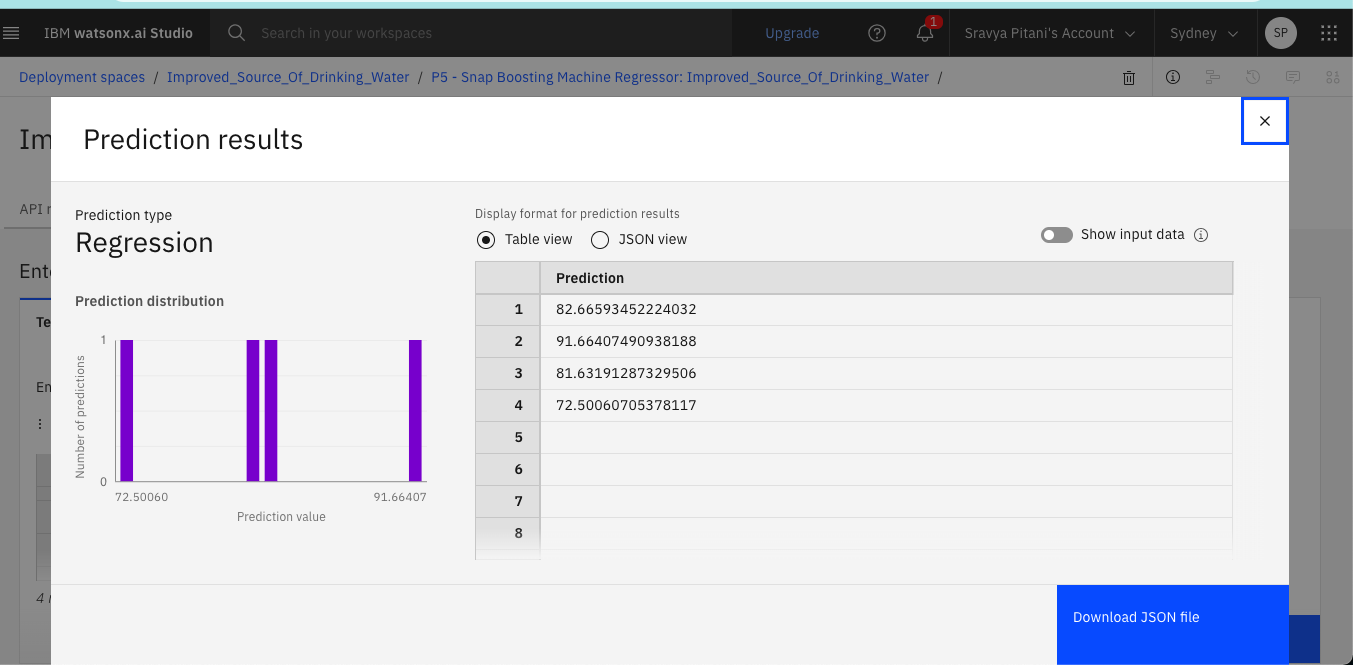
# REFERENCES

* AI Kosh Dataset – Improved Source of Drinking Water (78th Round Survey)

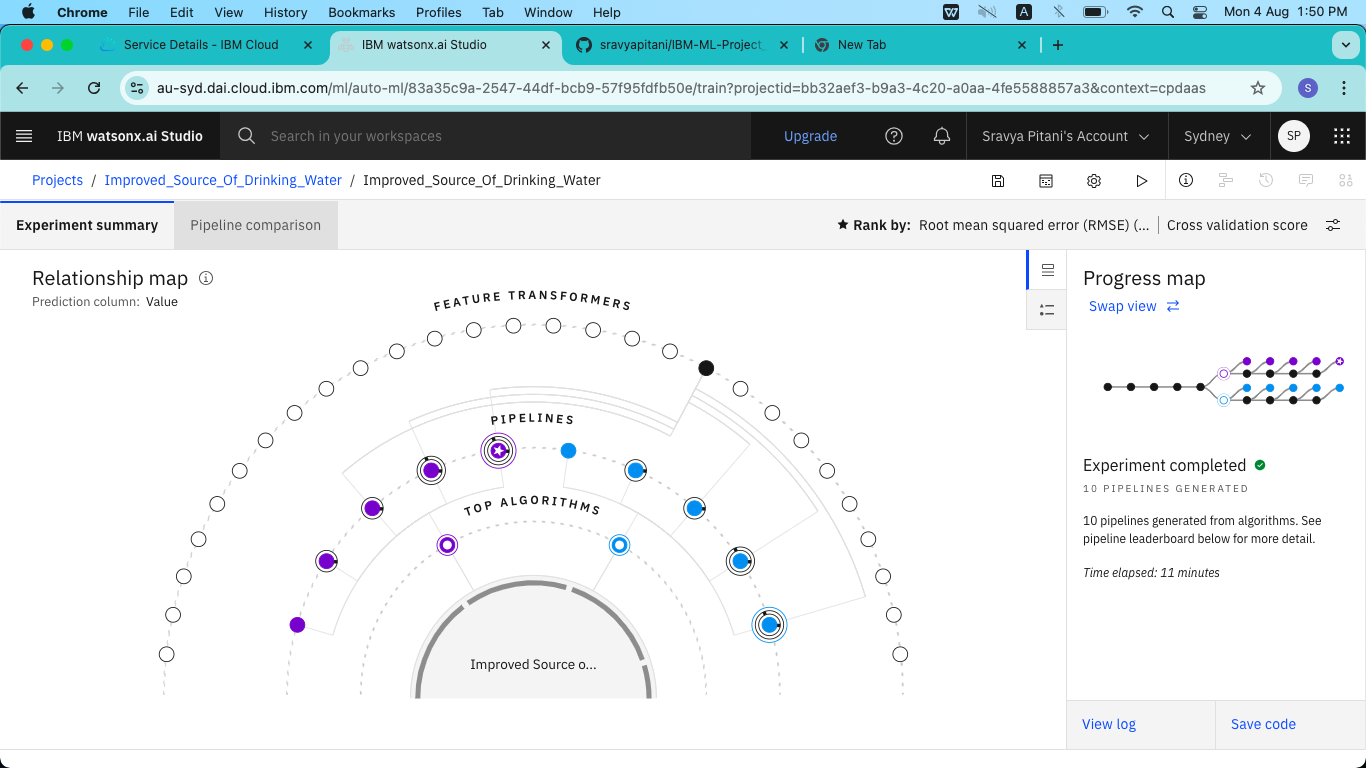
[https://aikosh.indiaai.gov.in/web/datasets/details/improved\_source\_of\_drinking\_wate](https://aikosh.indiaai.gov.in/web/datasets/details/improved_source_of_drinking_water_multiple_indicator_survey_78th_round.html) [r\_multiple\_indicator\_survey\_78th\_round.html](https://aikosh.indiaai.gov.in/web/datasets/details/improved_source_of_drinking_water_multiple_indicator_survey_78th_round.html)

* IBM Cloud Lite Documentation –Use of IBM Watson Studio, IBM Cloud Object Storage, and IBM Machine Learning Services – https://cloud.ibm.com
* Python Libraries – *Pandas, NumPy*

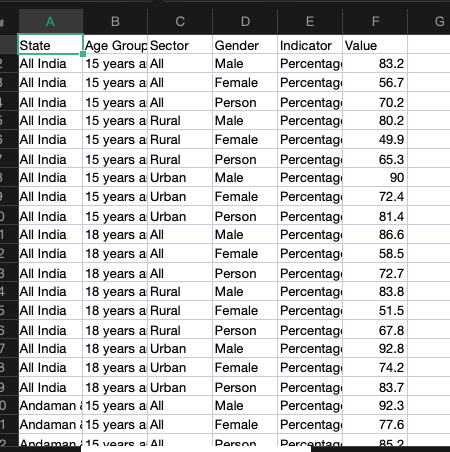
**SCREENSHOTS OF PROJECT**

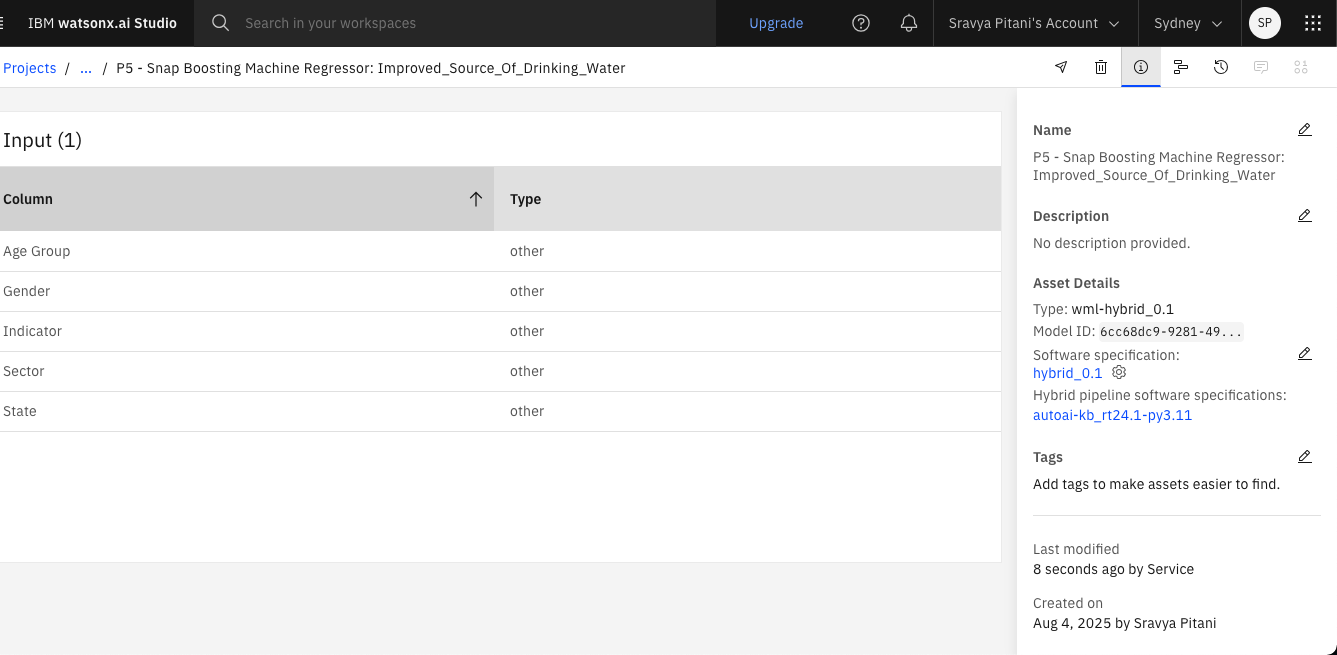
****

**SCREENSHOTS OF PROJECT**

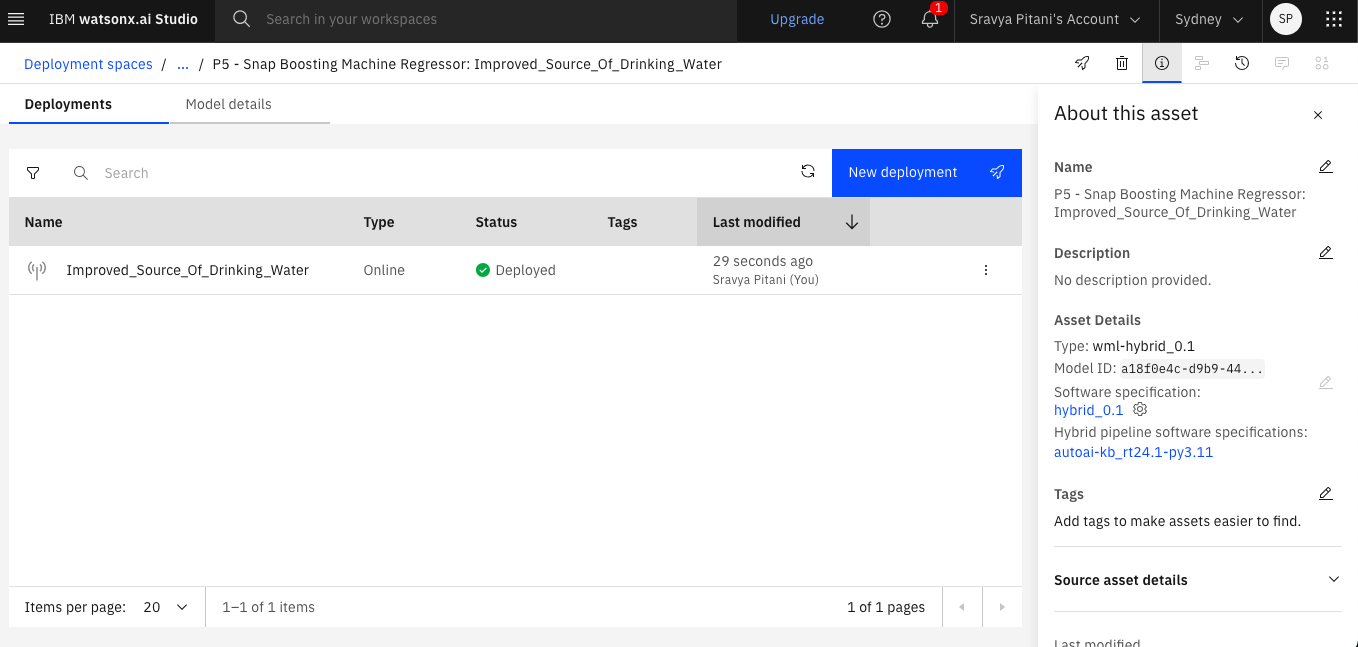
****

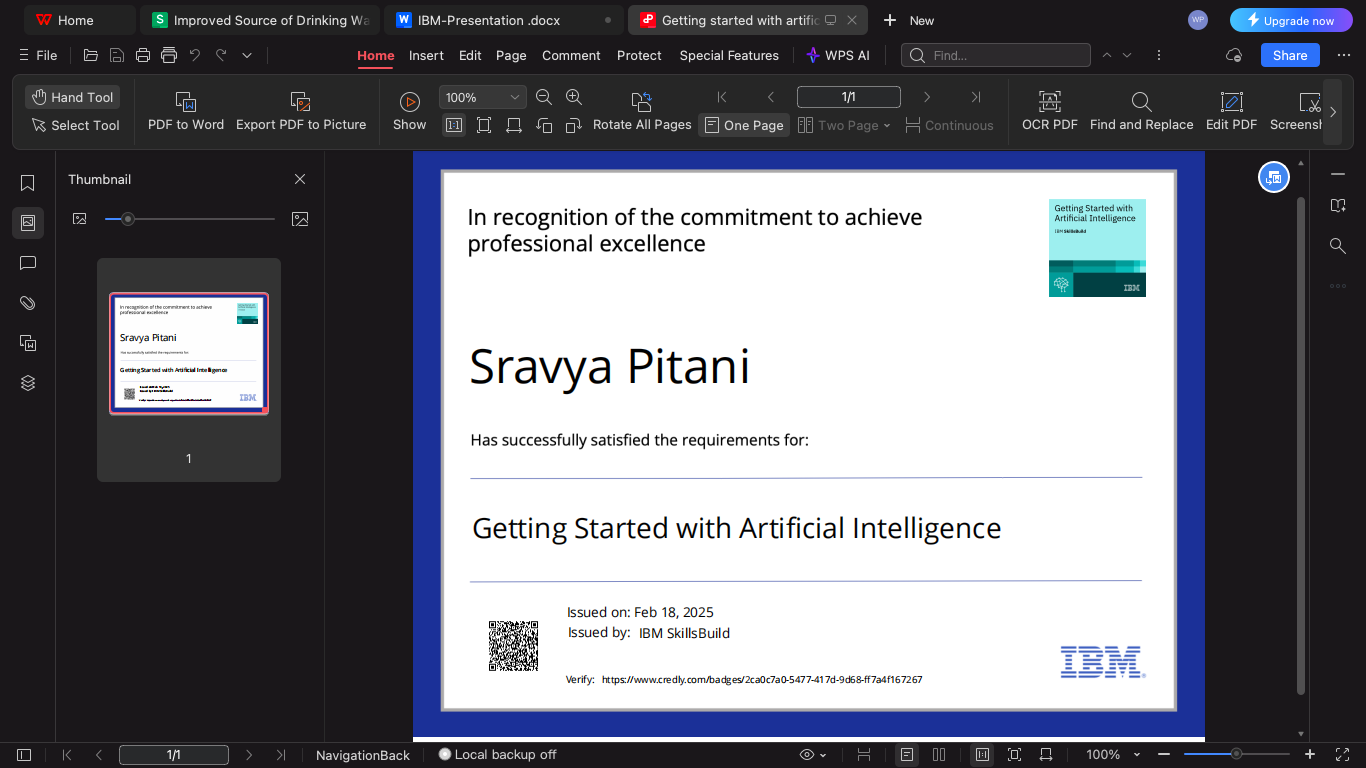
**SCREENSHOTS OF PROJECT**

****

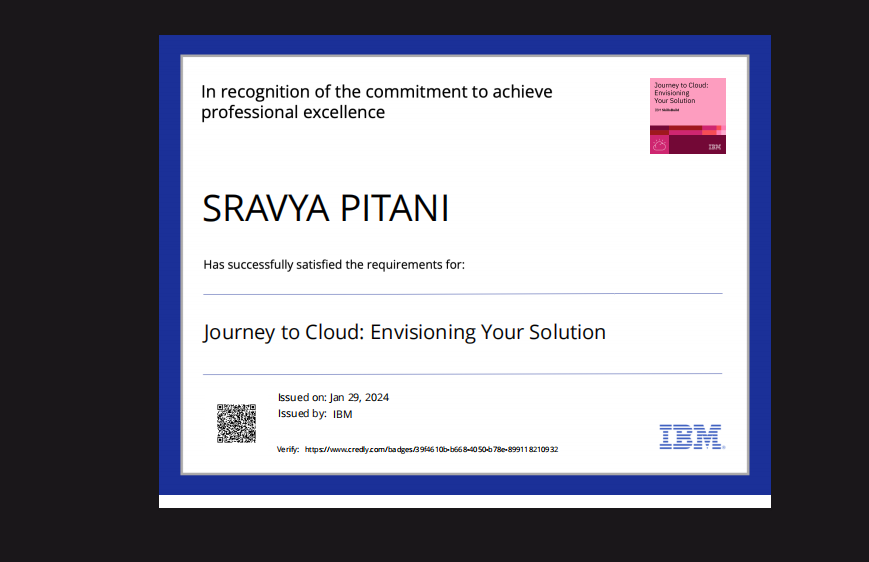
****

**SCREENSHOTS OF PROJECT**

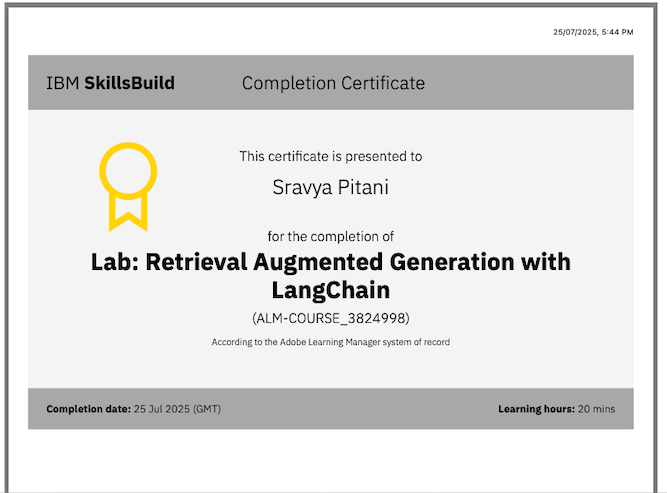
****

**IBM CERTIFICATIONS**

****

******IBM CERTIFICATIONS**

**IBM CERTIFICATIONS**

****

**THANK YOU**