CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS UNDER PMGSY USING MACHINE LEARNING ON IBM CLOUD

Presented By:
Bashaboina Renuka- Chaitanya Bharathi Institute of Technology(CBIT)- Computer Science and Engineering (Artificial Intelligence and Machine Learning)

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

The Pradhan Mantri Gram Sadak Yojana (PMGSY) is a rural development initiative aimed at improving road and bridge connectivity across India. The program includes multiple schemes like PMGSY-I, PMGSY-II, and RCPLWEA, each with distinct characteristics.

Manually classifying thousands of infrastructure projects into the correct scheme is time-consuming, error-prone, and not scalable.

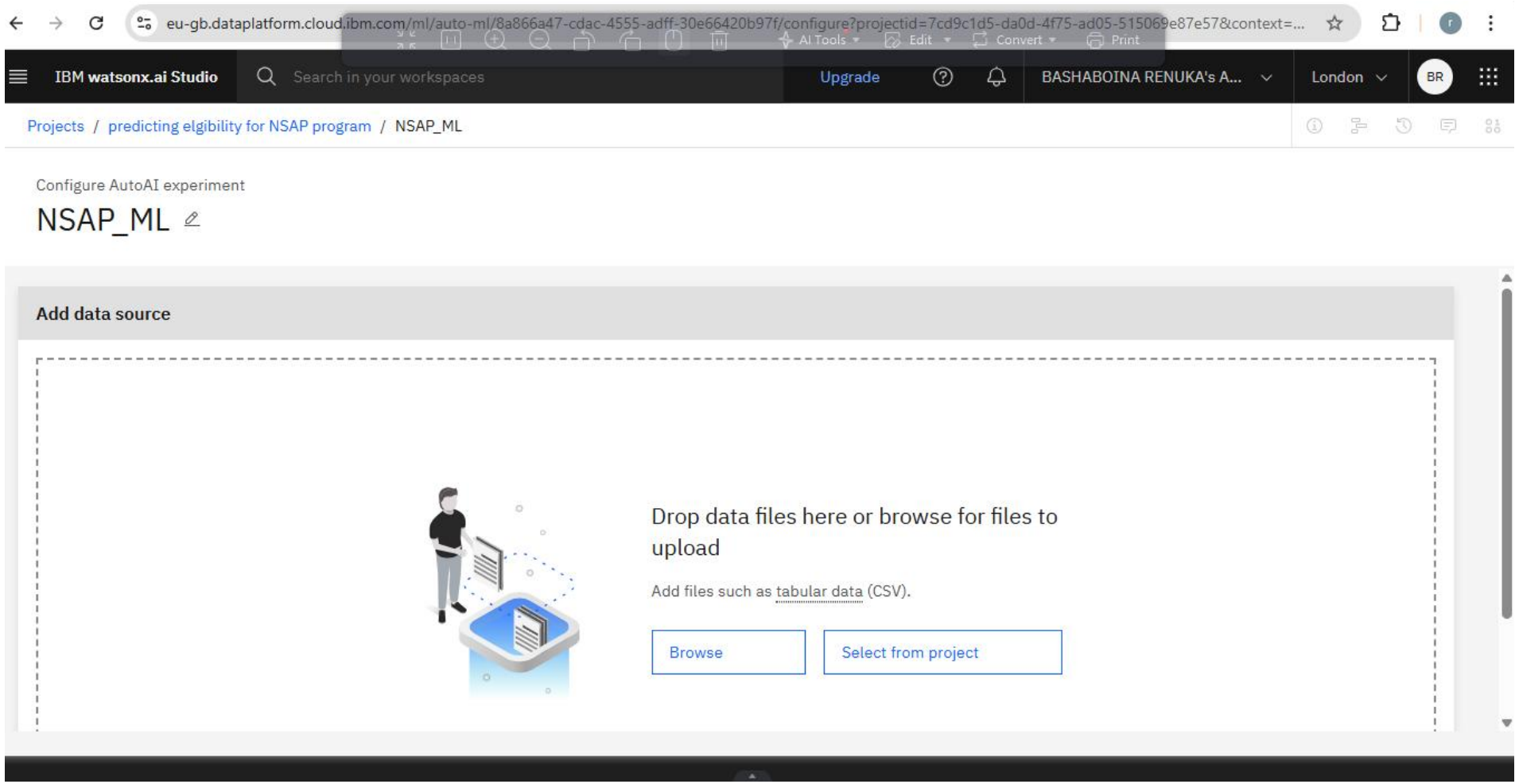
This challenge affects project monitoring, budget allocation, and policy planning.

An efficient and intelligent classification method is needed to streamline this process.

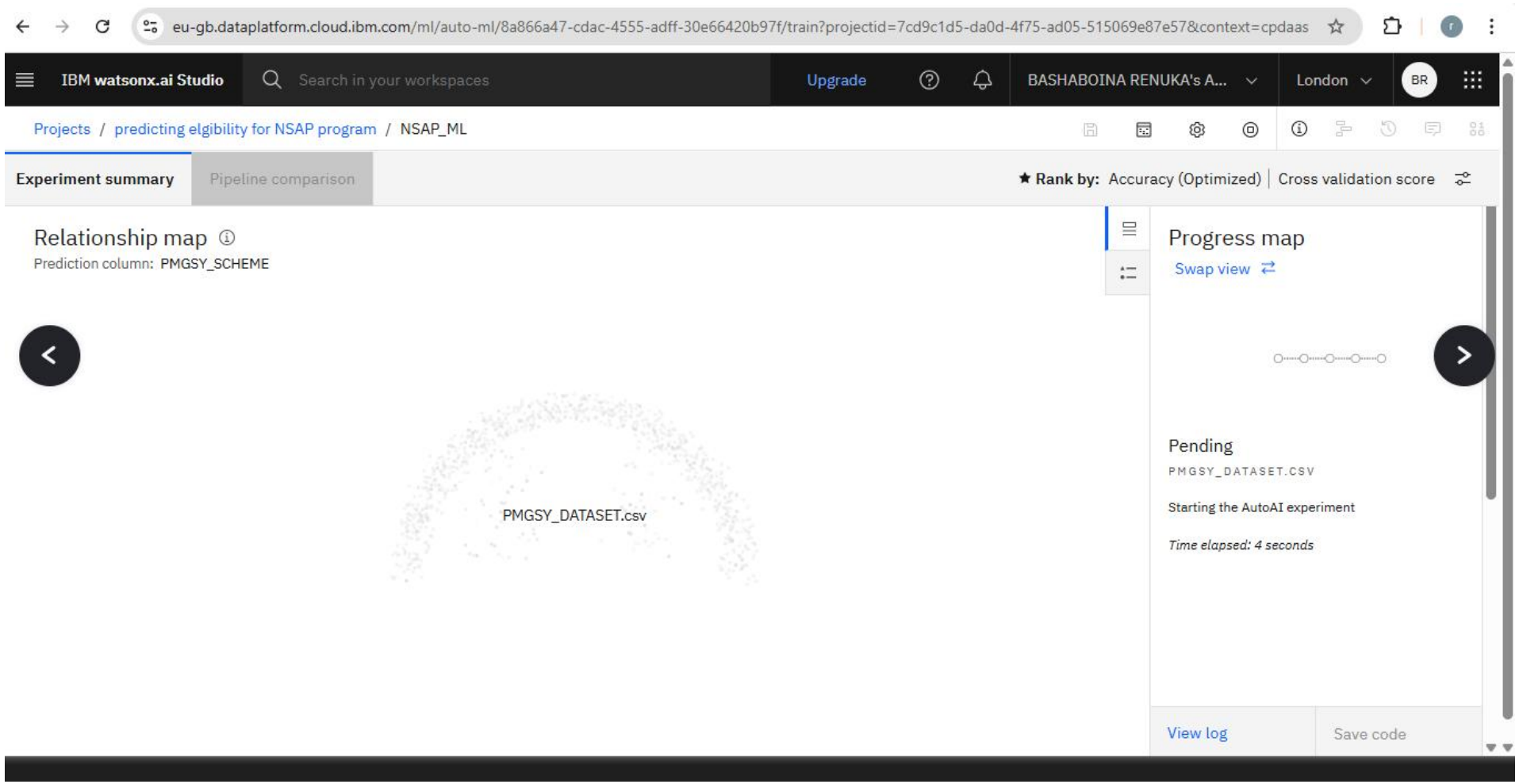
PROPOSED SOLUTION

- **The proposed system aims to address the challenge of automatically classifying rural infrastructure projects under their appropriate PMGSY schemes (PMGSY-I, PMGSY-II, RCPLWEA, etc.). This involves leveraging machine learning techniques and structured government data to streamline the classification process and reduce manual errors.The solution will consist of the following components:**
- **Data Collection:**
  - Gather and utilize historical project data provided through the AI-KOSH platform, which contains structured information such as state, district, road/bridge type, total cost, project length, and number of bridges or road segments.Gather and utilize historical project data provided through the AI-KOSH platform, which contains structured information such as state, district, road/bridge type, total cost, project length, and number of bridges or road segments.
- **Data Preprocessing:**
  - Prepare the dataset for model training by cleaning missing or inconsistent values, encoding categorical variables (e.g., project type, state), and selecting the most relevant features. Techniques such as one-hot encoding and normalization may be applied where necessary.
- **Machine Learning Algorithm:**
  - Implement a multi-class classification model, such as a Random Forest Classifier or use IBM's AutoAI tool, which automatically selects and tunes the best model. The model will learn from historical project data to predict the most appropriate scheme for new or unlabelled projects.
- **Deployment:**
  - The trained model will be deployed using IBM Watson Machine Learning, exposing a RESTful API that can be integrated into other government dashboards or used for batch processing of new project data.
- **Evaluation:**
  - Evaluate the model's accuracy, precision, recall, and F1-score using test data. The model will be fine-tuned and monitored periodically to ensure consistent performance and adaptability to new data patterns.

SELECTING DATA



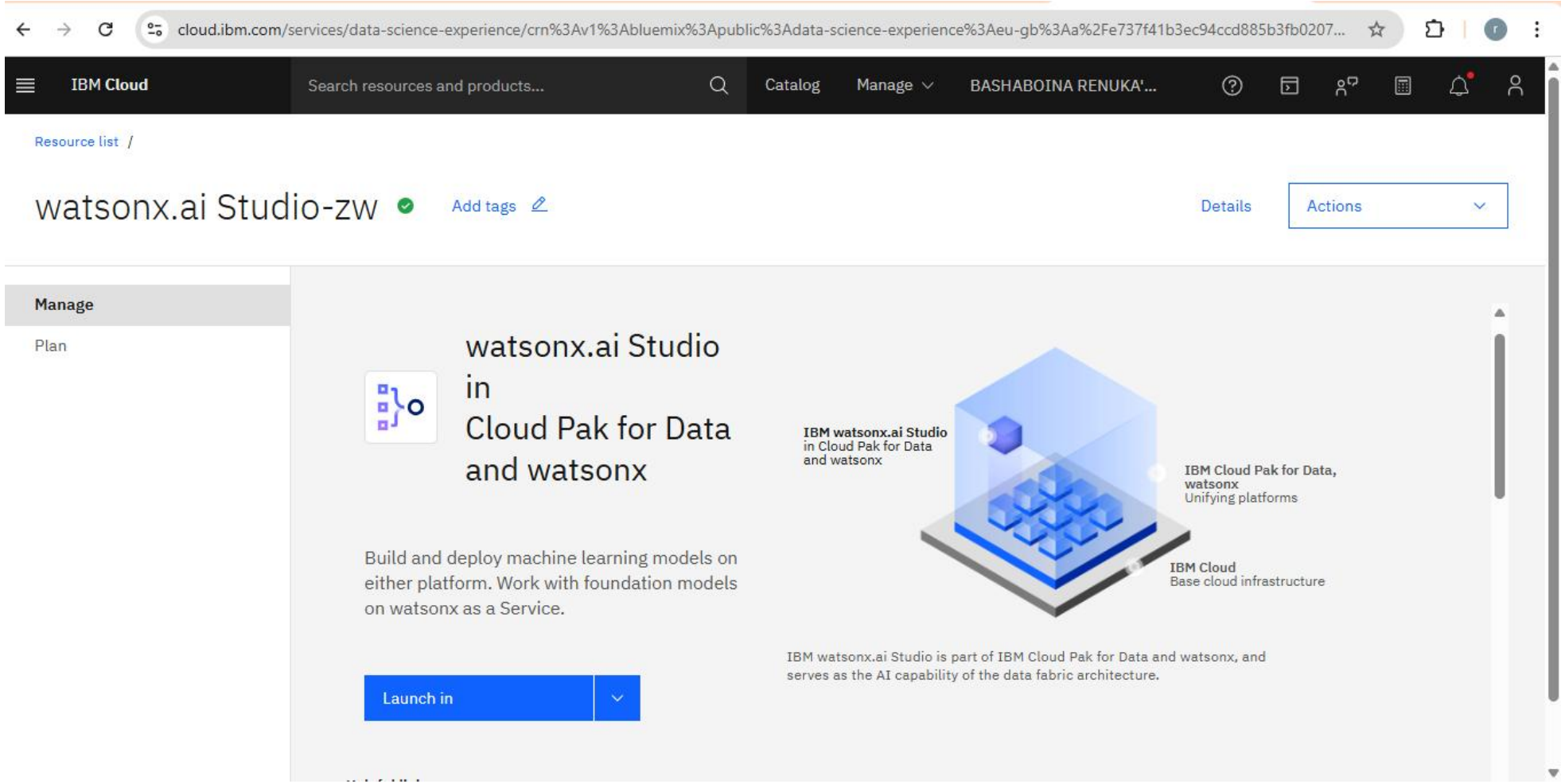
UPLOADED DATASET



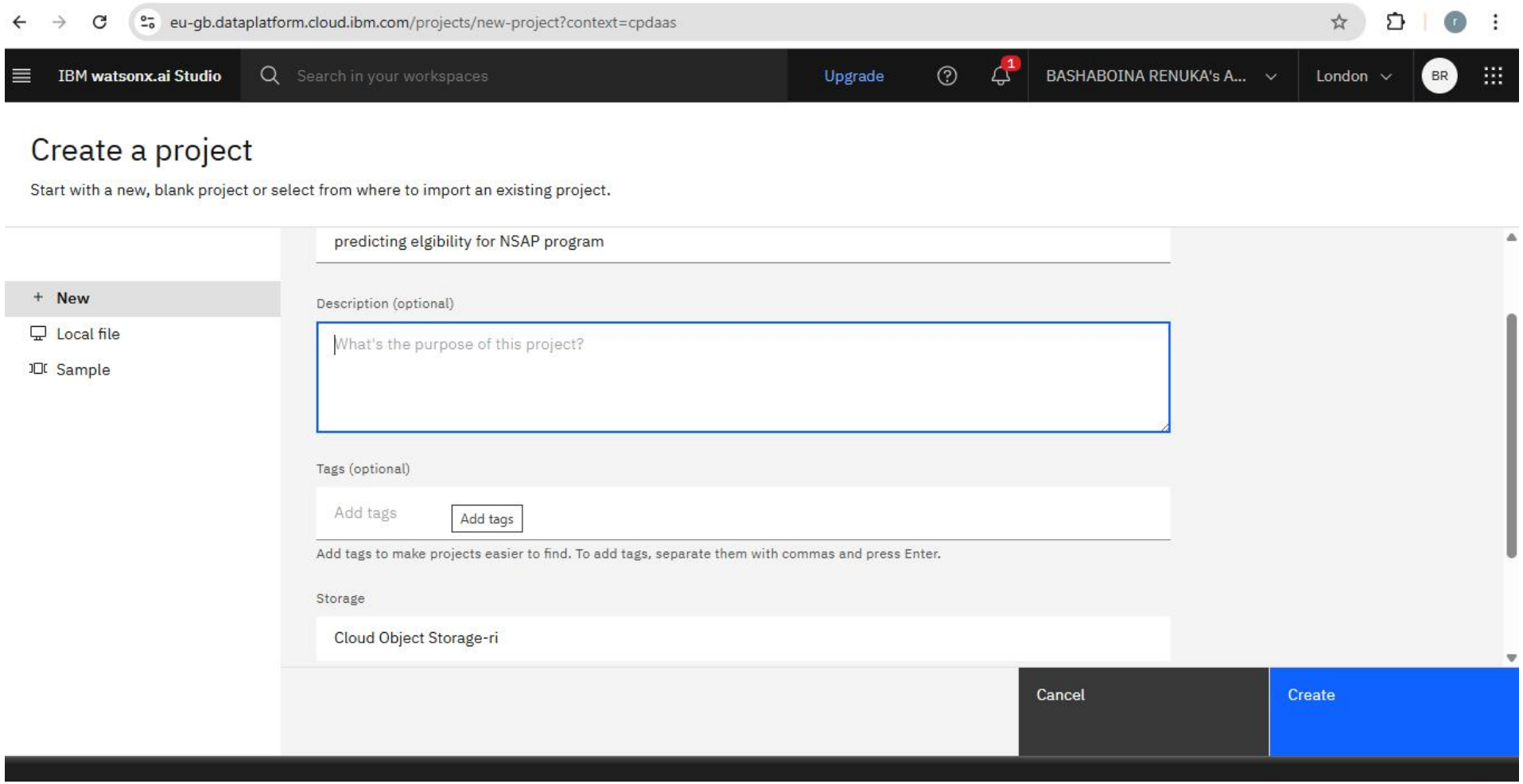
SYSTEM APPROACH

- The System Approach section outlines the overall strategy and methodology for developing and implementing the PMGSY scheme classification model. The solution is built using IBM Cloud services along with standard data science tools and machine learning libraries.
- System requirements:
  - IBM Cloud Lite Account (Free tier)
  - Watson Studio – for data analysis, notebook execution, and AutoAI
  - Cloud Object Storage – to store datasets
  - Watson Machine Learning – for model deployment
- Library required to build the model:
  - pandas – For data loading, preprocessing, and manipulation
  - numpy – For numerical computations and array handling
  - scikit-learn –
  - For machine learning algorithms (e.g., RandomForestClassifier)
  - For model evaluation (accuracy, precision, recall, etc.)
  - matplotlib / seaborn – For data visualization (optional but useful for EDA)
  - joblib – For saving/training the ML model (model serialization)

LAUNCHING WATSONX.AI STUDIO

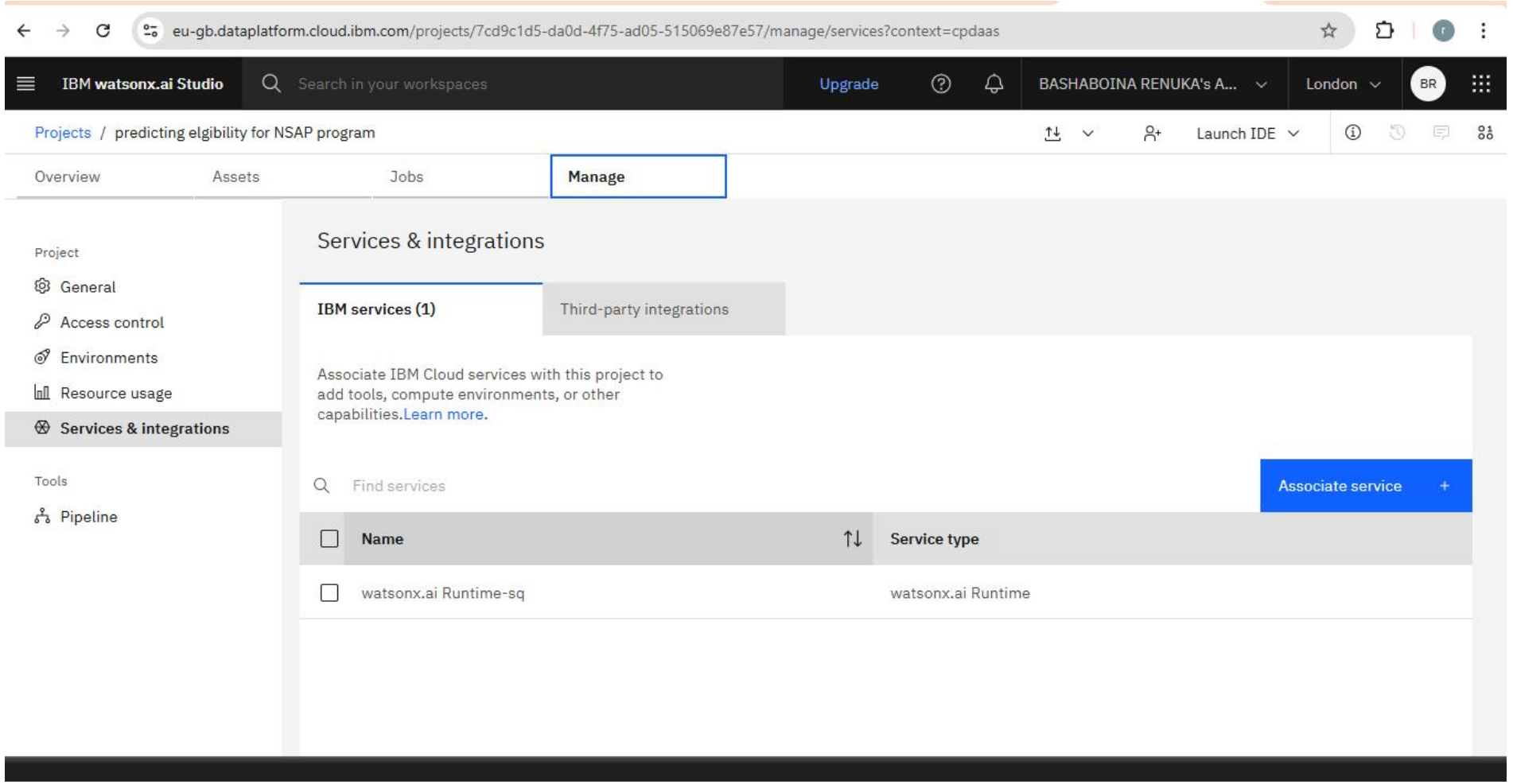


TITLE OF PROJECT



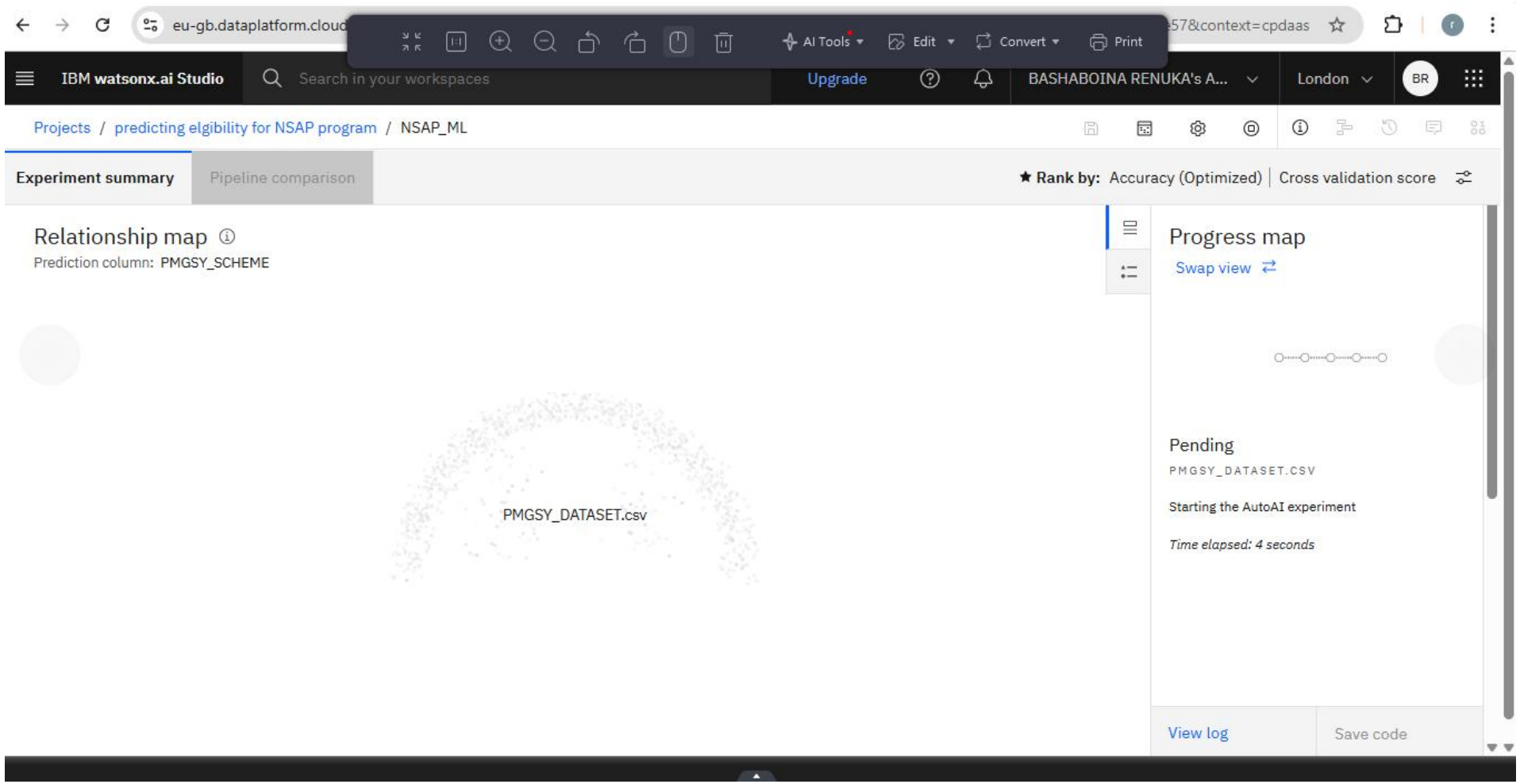
edunet foundation

ASSOCIATE SERVICES

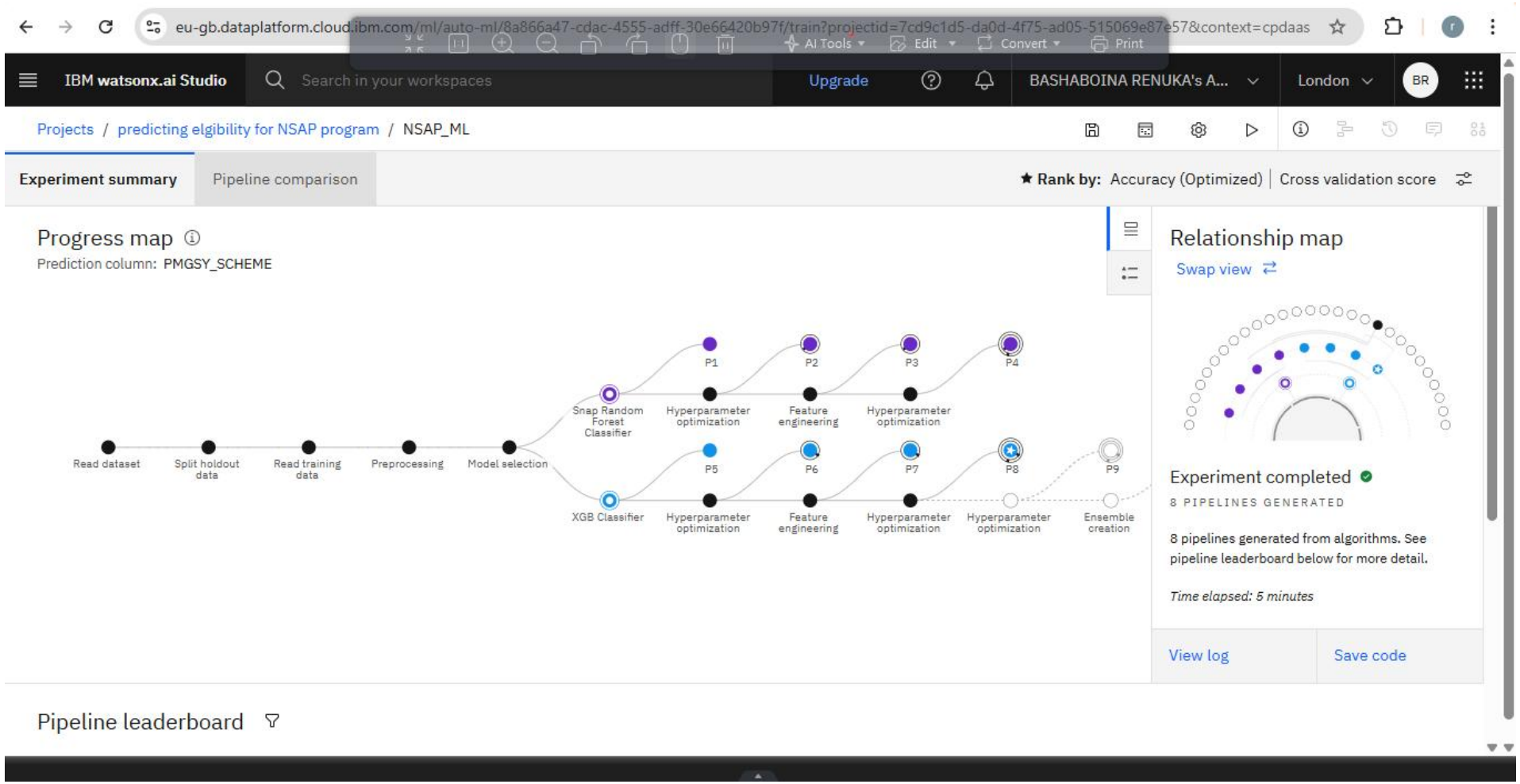


edunet foundation

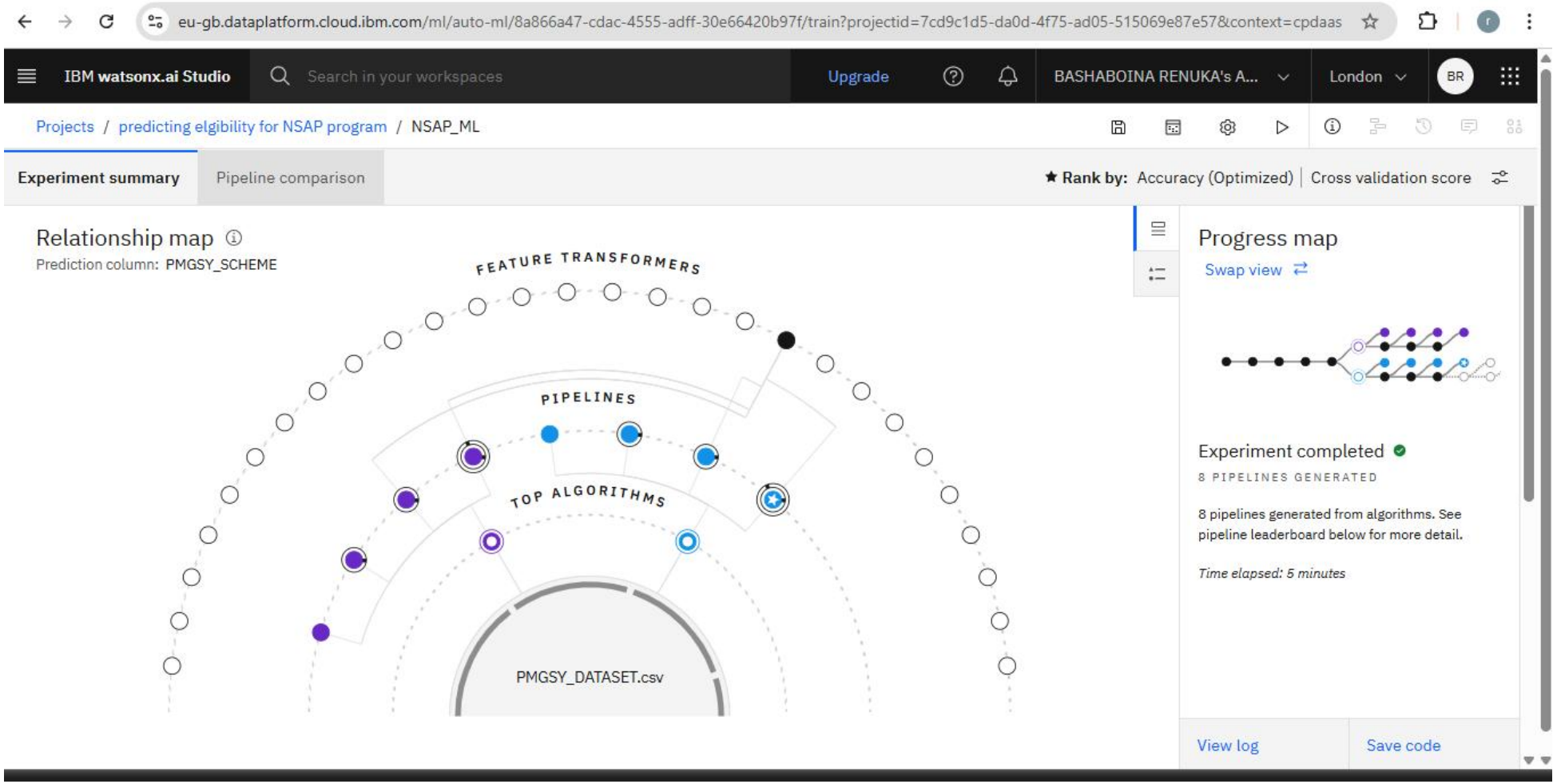
BUILDING MODEL



PROCESSING DATA

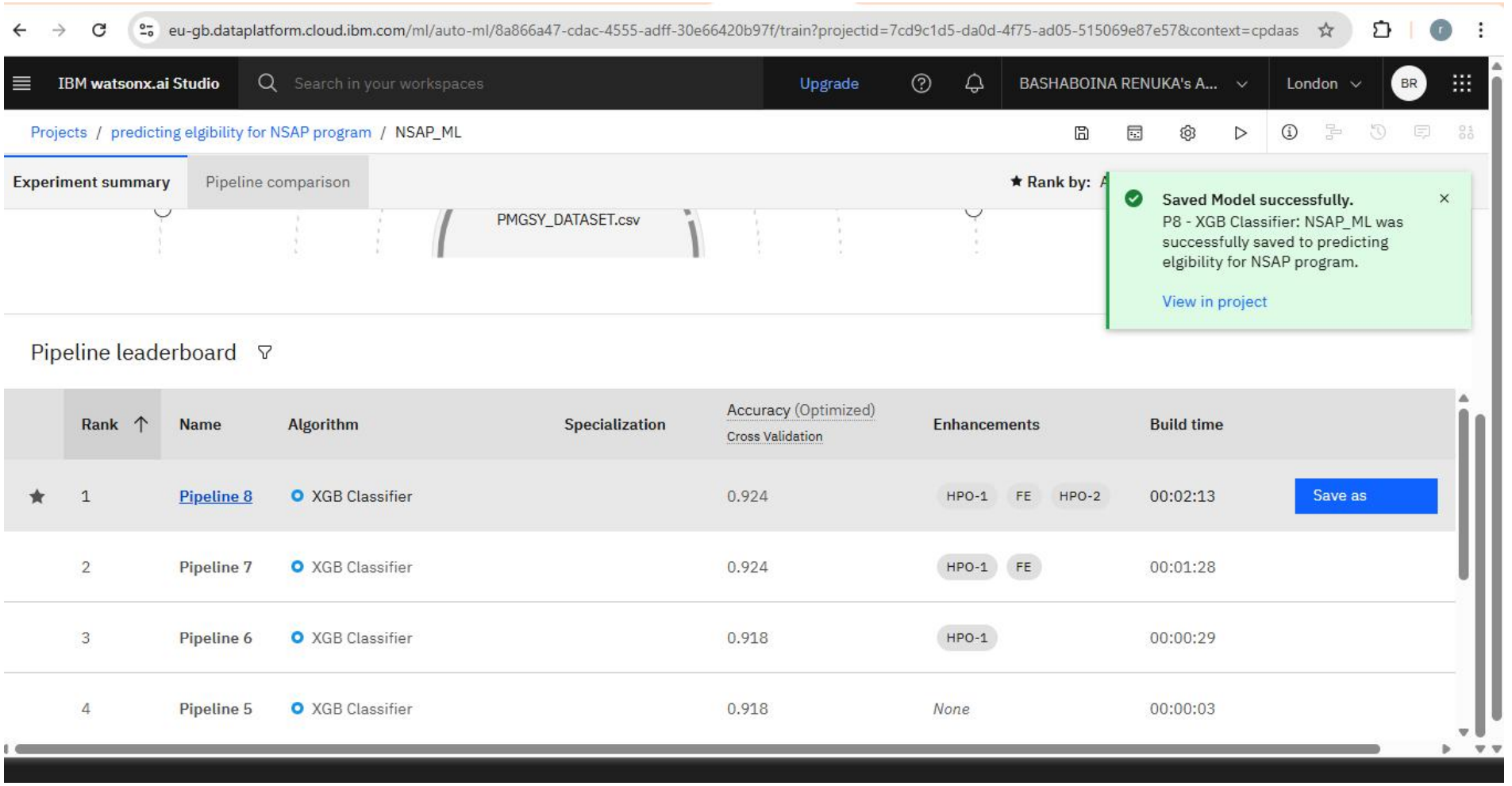


SELECTING BEST ALGORITHM



edunet foundation

SELECTING BEST PIPELINE

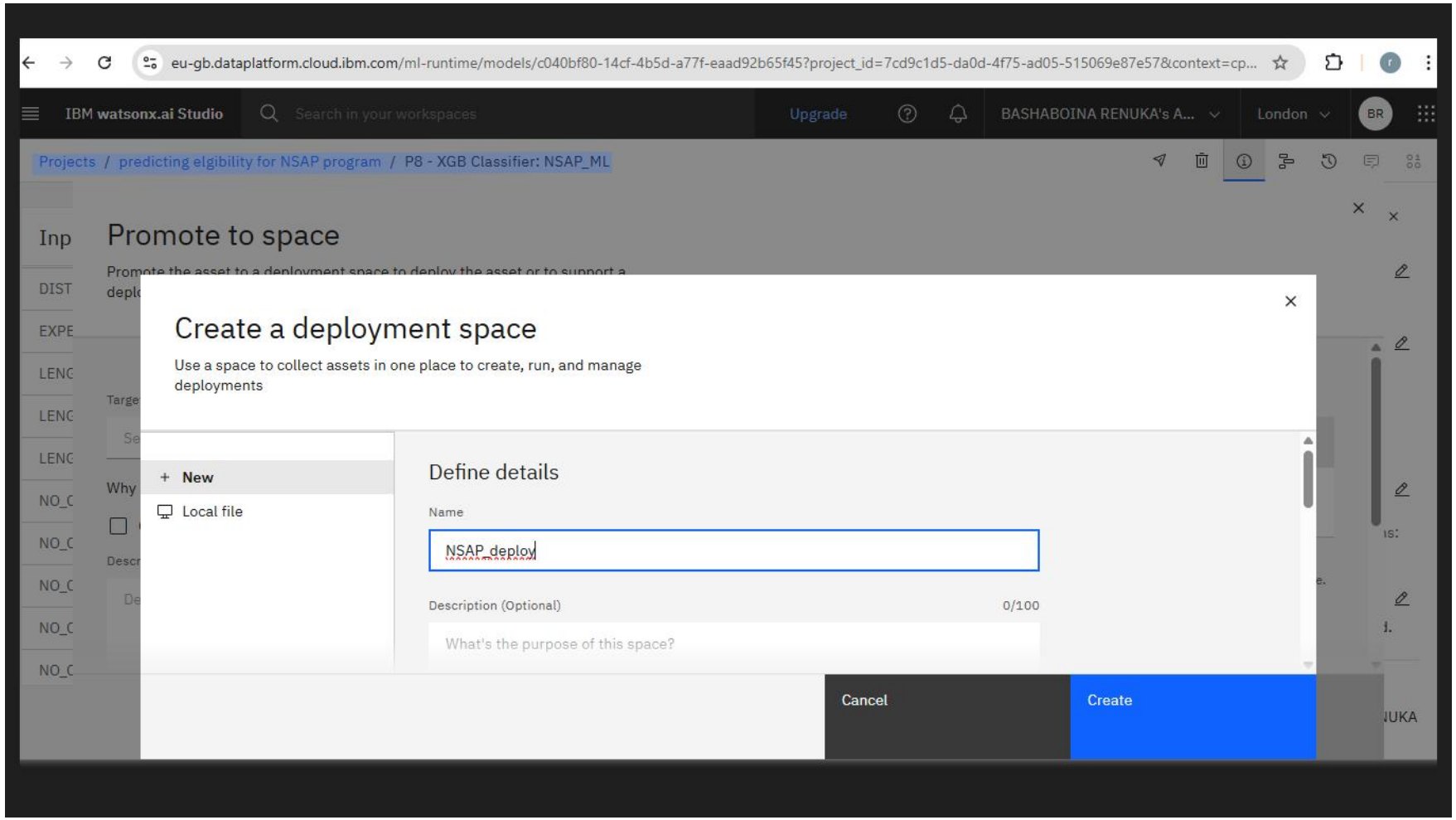


edunet foundation

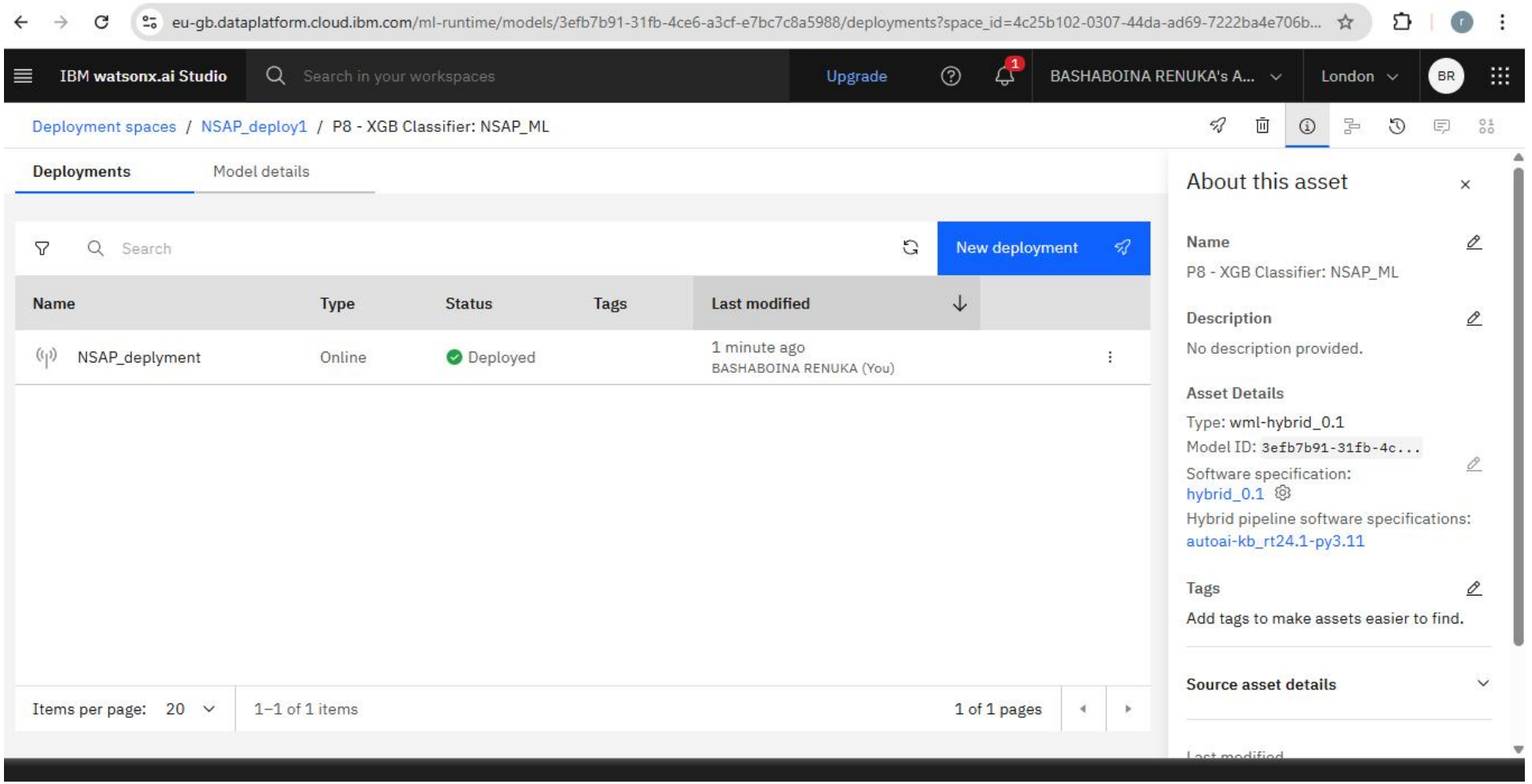
ALGORITHM & DEPLOYMENT

- In the Algorithm section, we describe the machine learning algorithm chosen for classifying rural infrastructure projects under their respective PMGSY schemes.
- Algorithm Selection:
  - Provide a brief overview of the chosen algorithm (e.g., time-series forecasting model, like ARIMA or LSTM) and justify its selection based on the problem statement and data characteristics.
- Data Input:
  - Specify the input features used by the algorithm, such as historical bike rental data, weather conditions, day of the week, and any other relevant factors.
- Training Process:
  - Explain how the algorithm is trained using historical data. Highlight any specific considerations or techniques employed, such as cross-validation or hyperparameter tuning.
- Prediction Process:
  - Detail how the trained algorithm makes predictions for future bike counts. Discuss any real-time data inputs considered during the prediction phase.

DEPLOYMENT SPACE

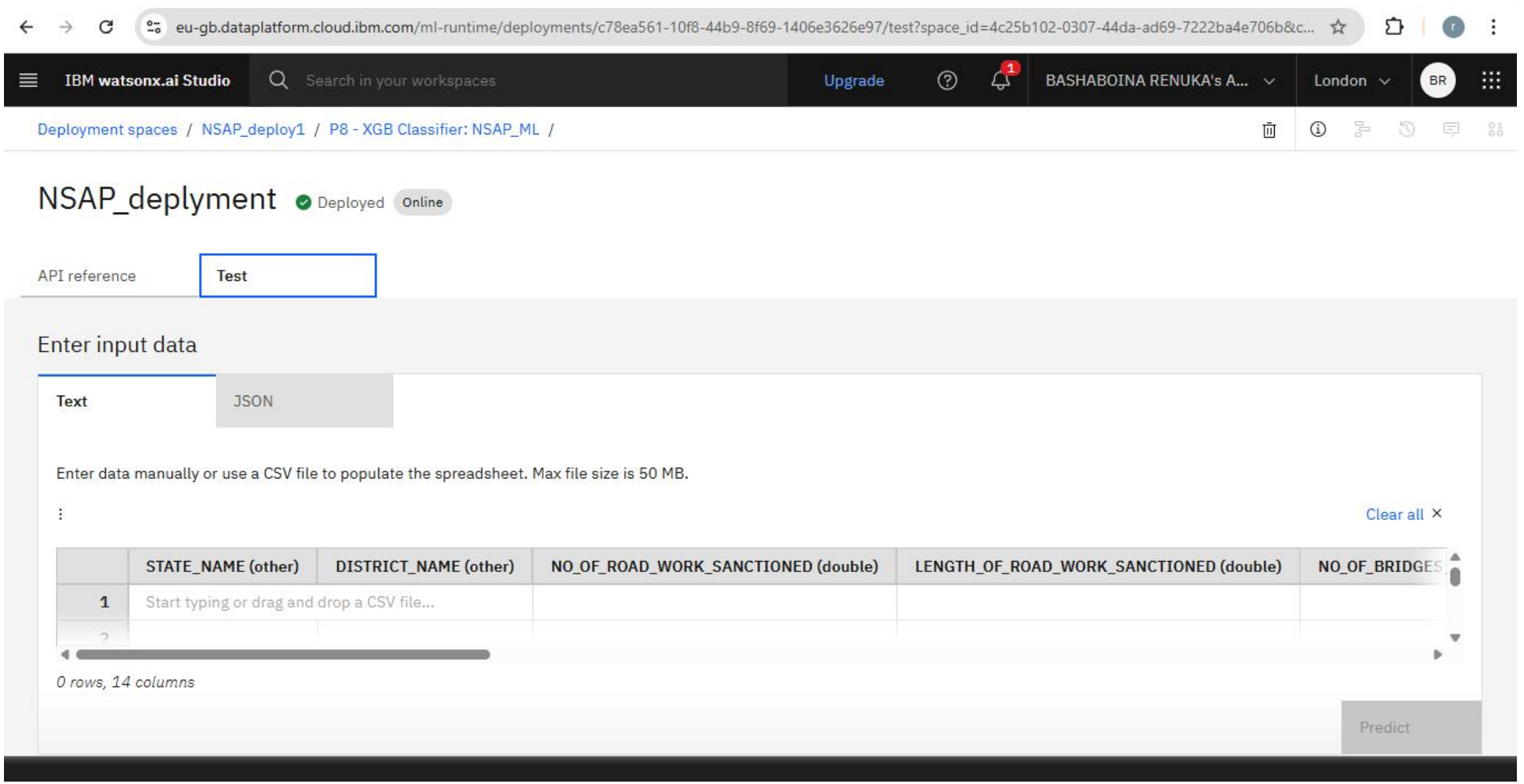


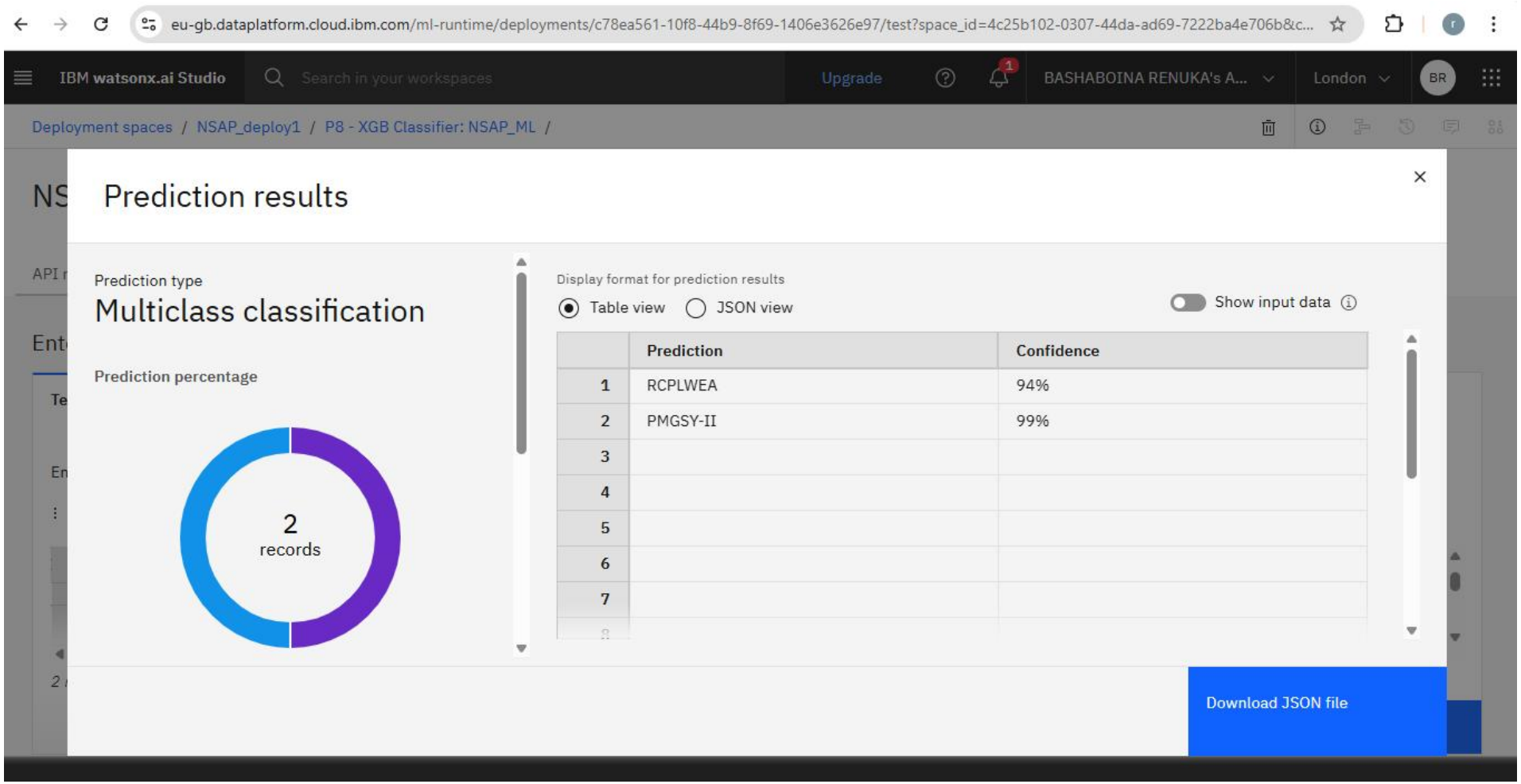
edunet foundation



RESULT

The machine learning model was trained to classify rural infrastructure projects into appropriate PMGSY schemes based on features such as cost, length, number of bridges, and state information. The model's performance was evaluated using standard classification metrics.





CONCLUSION

- In this project, we developed a machine learning model to automatically classify rural infrastructure projects under the appropriate PMGSY scheme (PMGSY-I, PMGSY-II, RCPLWEA, etc.). By using structured data from the AI-KOSH dataset and applying classification algorithms such as Random Forest, we were able to build a system that reduces manual effort and improves accuracy in scheme identification.

- The project demonstrated the potential of using AI to support government planning and monitoring processes. With features like project cost, road length, and number of bridges, the model successfully predicted the scheme with good accuracy. The model was deployed using IBM Watson Machine Learning, allowing real-time predictions through an API.

- This system can save time, reduce human errors, and help decision-makers better allocate resources. Overall, the project shows how data and AI can be effectively used in rural development planning and governance.

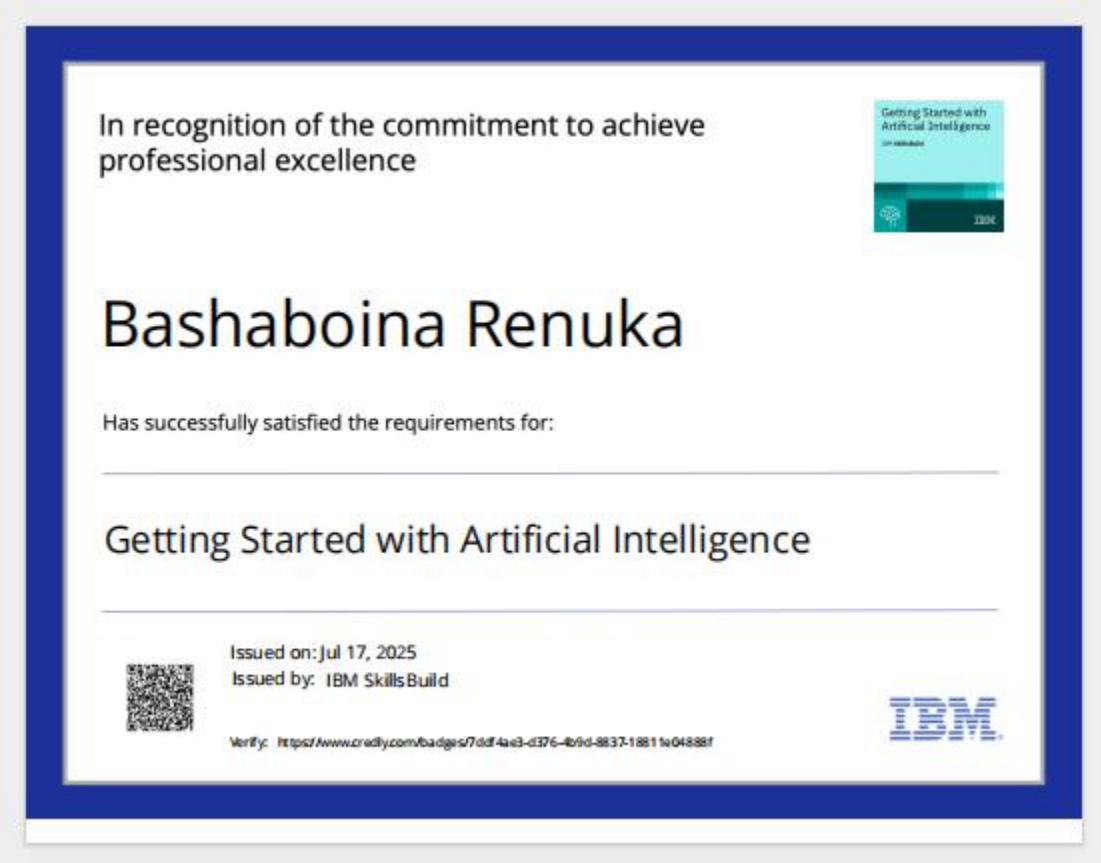
FUTURE SCOPE

- The current system effectively classifies rural infrastructure projects under appropriate PMGSY schemes based on existing physical and financial characteristics. However, there are several ways this system can be further enhanced and expanded:
- Adding More Data
- Right now, your model uses basic project info like cost and road length.
- In the future, we can add:
- Start and end dates of the project
- Weather or terrain (hill, forest, etc.)
- Delays or contractor quality
- This will help the model give more accurate predictions.

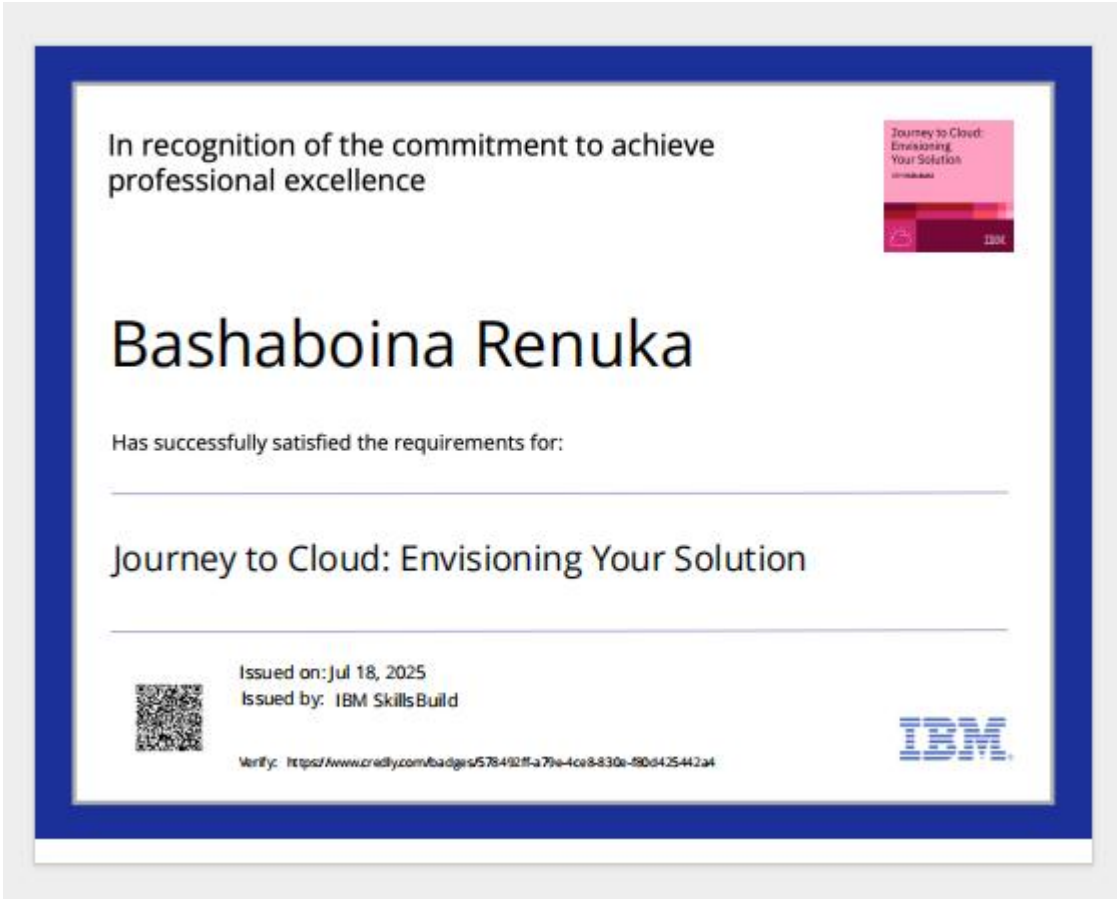
REFERENCES

- AI-KOSH Dataset – PMGSY Project Data
- Government of India, National AI Portal
- URL: https://aikosh.indiaai.gov.in
- Scikit-learn: Machine Learning in Python
- Pedregosa et al., Journal of Machine Learning Research, 2011
- URL: https://scikit-learn.org/
- A key library used for classification models like Random Forest, Decision Trees, etc.
- IBM Watson Studio Documentation
- IBM Cloud Docs
- URL: https://dataplatform.cloud.ibm.com/docs
- Random Forests
- Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
- Foundational paper explaining how Random Forest classifiers work.
- A Survey on Classification Techniques in Data Mining
- S. Kotsiantis, Informatica, 2007
- URL: https://www.informatica.si/PDF/31-3/05_Kotsiantis%20-%20Classification%20techniques.pdf
- Useful for understanding and comparing various classification algorithms.
- Data Preprocessing Techniques in Machine Learning
- Dhanalakshmi et al., International Journal of Computer Applications, 2016
- Covers standard techniques for cleaning and preparing structured datasets.

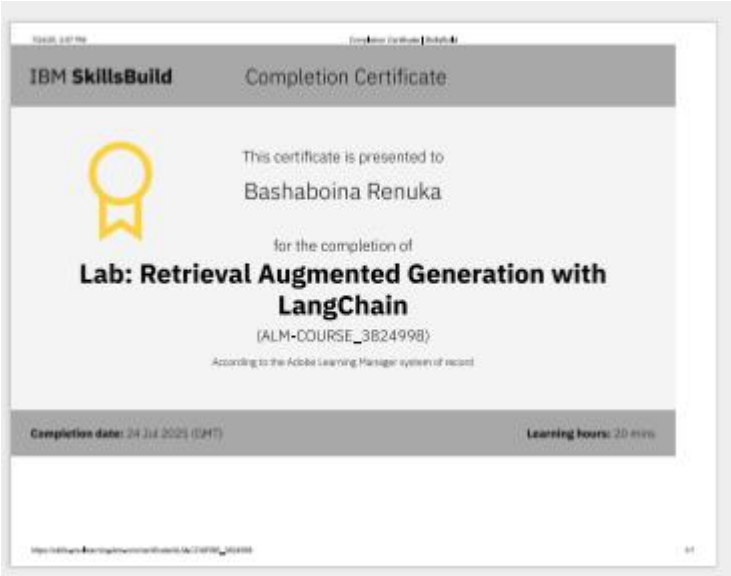
IBM CERTIFICATIONS



IBM CERTIFICATIONS



IBM CERTIFICATIONS


IBM SkillsBuild

Completion Certificate

This certificate is presented to

Bashaboina Renuka

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