

---

# **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)**

# 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.
- **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

eu-gb.dataplatform.cloud.ibm.com/ml/auto-ml/8a866a47-cdac-4555-adff-30e66420b97f/configure?projectid=7cd9c1d5-da0d-4f75-ad05-515069e87e57&context=...

IBM watsonx.ai Studio Search in your workspaces Upgrade ? BASHABOINA RENUKA's A... London BR

Projects / predicting eligibility for NSAP program / NSAP\_ML

Configure AutoAI experiment  
NSAP\_ML

Add data source

Drop data files here or browse for files to upload

Add files such as tabular data (CSV).

Browse Select from project

# UPLOADED DATASET

The screenshot displays the IBM Watsonx.ai Studio web interface. The browser address bar shows the URL: `eu-gb.dataplatform.cloud.ibm.com/ml/auto-ml/8a866a47-cdac-4555-adff-30e66420b97f/train?projectId=7cd9c1d5-da0d-4f75-ad05-515069e87e57&context=cpdaas`. The top navigation bar includes the IBM Watsonx.ai Studio logo, a search bar, an 'Upgrade' button, and user information for 'BASHABOINA RENUKA's A...'. The breadcrumb trail indicates the current location: 'Projects / predicting eligibility for NSAP program / NSAP\_ML'. The main interface is divided into two tabs: 'Experiment summary' (active) and 'Pipeline comparison'. The 'Experiment summary' tab shows a 'Relationship map' with the prediction column 'PMGSY\_SCHEME' and a 'Progress map' on the right. The 'Progress map' shows the status 'Pending' for the dataset 'PMGSY\_DATASET.CSV', with the message 'Starting the AutoAI experiment' and 'Time elapsed: 4 seconds'. The 'View log' and 'Save code' buttons are visible at the bottom of the progress map. The background of the 'Relationship map' features a large, faint watermark of the IBM logo.

# 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


cloud.ibm.com/services/data-science-experience/crn%3Av1%3Abluemix%3Apublic%3Adata-science-experience%3Aeu-gb%3Aa%2Fe737f41b3ec94ccd885b3fb0207...

IBM Cloud Search resources and products... Catalog Manage BASHABOINA RENUKA'...

Resource list /


watsonx.ai Studio-zw ✓ Add tags [Details](#) [Actions](#)

Manage Plan

 **watsonx.ai Studio**  
in  
**Cloud Pak for Data**  
and **watsonx**

Build and deploy machine learning models on either platform. Work with foundation models on watsonx as a Service.

[Launch in](#) [▼](#)



IBM watsonx.ai Studio in Cloud Pak for Data and watsonx

IBM Cloud Pak for Data, watsonx Unifying platforms

IBM Cloud Base cloud infrastructure

IBM watsonx.ai Studio is part of IBM Cloud Pak for Data and watsonx, and serves as the AI capability of the data fabric architecture.



# TITLE OF PROJECT

←

→

↺

eu-gb.dataplatform.cloud.ibm.com/projects/new-project?context=cpdaas

☆

🔖

👤

⋮

☰ IBM watsonx.ai Studio

🔍 Search in your workspaces

Upgrade

?

🔔 1

BASHABOINA RENUKA's A... ▾

London ▾

BR

⋮

Create a project

Start with a new, blank project or select from where to import an existing project.

+ New

📁 Local file

📁 Sample

predicting eligibility for NSAP program

Description (optional)

What's the purpose of this project?

Tags (optional)

Add tags

Add tags

Add tags to make projects easier to find. To add tags, separate them with commas and press Enter.

Storage

Cloud Object Storage-ri

Cancel

Create

# ASSOCIATE SERVICES

The screenshot shows the IBM watsonx.ai Studio interface. The browser address bar displays the URL: eu-gb.dataplatform.cloud.ibm.com/projects/7cd9c1d5-da0d-4f75-ad05-515069e87e57/manage/services?context=cpdaas. The top navigation bar includes the IBM watsonx.ai Studio logo, a search bar, an 'Upgrade' button, and user information for BASHABOINA RENUKA's A... in the London region. The main breadcrumb is 'Projects / predicting eligibility for NSAP program'. The 'Manage' tab is selected, showing a sub-tab for 'Services & integrations'. The left sidebar lists project settings: General, Access control, Environments, Resource usage, and Services & integrations (which is highlighted). The 'Services & integrations' section has two tabs: 'IBM services (1)' and 'Third-party integrations'. A text block explains that IBM Cloud services can be associated with the project to add tools, compute environments, or other capabilities, with a 'Learn more' link. Below this is a search bar labeled 'Find services' and a blue 'Associate service +' button. A table lists the associated services:

<input type="checkbox"/>	Name	Service type
<input type="checkbox"/>	watsonx.ai Runtime-sq	watsonx.ai Runtime

# BUILDING MODEL

The screenshot displays the IBM watsonx.ai Studio web interface. The browser address bar shows the URL `eu-gb.dataplatform.cloud`. The top navigation bar includes the IBM watsonx.ai Studio logo, a search bar, an 'Upgrade' button, and user information for 'BASHABOINA RENUKA's A...'. The main content area is titled 'Projects / predicting eligibility for NSAP program / NSAP\_ML'. Below this, there are two tabs: 'Experiment summary' (selected) and 'Pipeline comparison'. The 'Experiment summary' tab shows a 'Relationship map' with the prediction column 'PMGSY\_SCHEME' and a 'Progress map' on the right. The 'Progress map' shows a 'Pending' status for 'PMGSY\_DATASET.CSV', with the message 'Starting the AutoAI experiment' and 'Time elapsed: 4 seconds'. At the bottom of the progress map, there are buttons for 'View log' and 'Save code'.

eu-gb.dataplatform.cloud

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

BASHABOINA RENUKA's A...

London

BR

Projects / predicting eligibility for NSAP program / NSAP\_ML

Experiment summary

Pipeline comparison

★ Rank by: Accuracy (Optimized) | Cross validation score

Relationship map ⓘ

Prediction column: PMGSY\_SCHEME

PMGSY\_DATASET.csv

Progress map

Swap view ↔

Pending

PMGSY\_DATASET.CSV

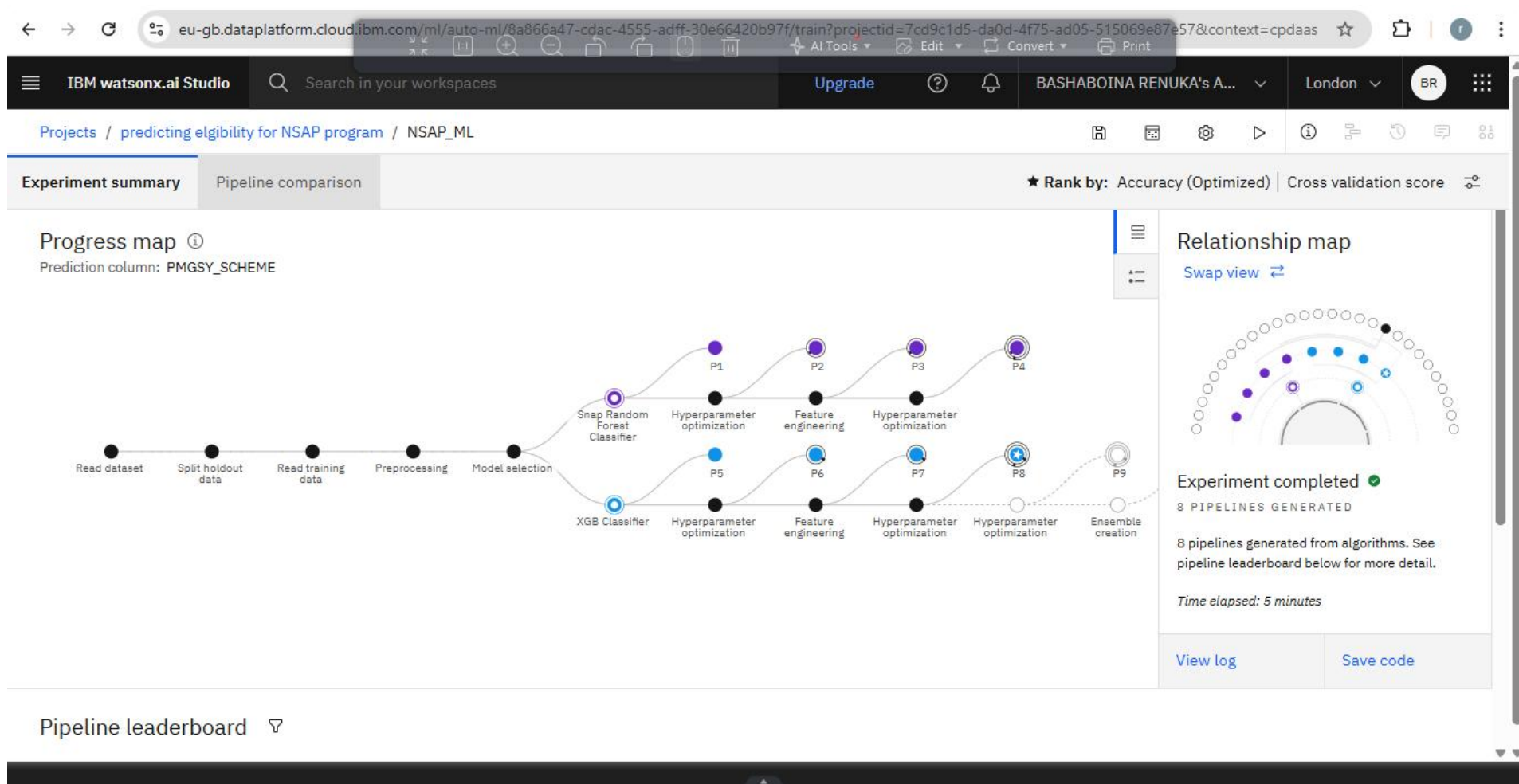
Starting the AutoAI experiment

Time elapsed: 4 seconds

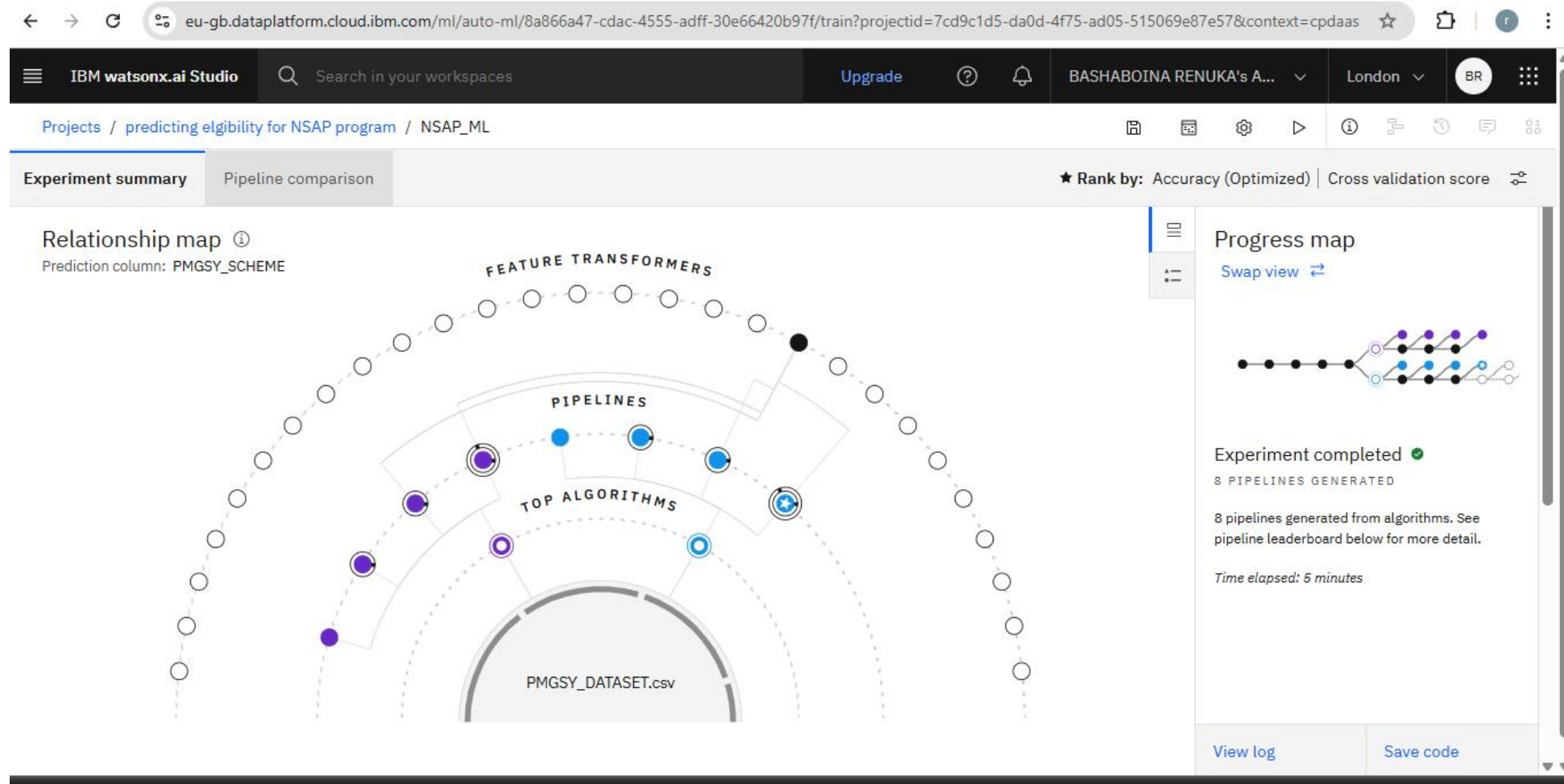
View log

Save code

# PROCESSING DATA



# SELECTING BEST ALGORITHM



# SELECTING BEST PIPELINE

eu-gb.dataplatform.cloud.ibm.com/ml/auto-ml/8a866a47-cdac-4555-adff-30e66420b97f/train?projectid=7cd9c1d5-da0d-4f75-ad05-515069e87e57&context=cpdaas

IBM watsonx.ai Studio Search in your workspaces Upgrade ? BASHABOINA RENUKA's A... London BR

Projects / predicting eligibility for NSAP program / NSAP\_ML

Experiment summary Pipeline comparison ★ Rank by: A

PMGSY\_DATASET.csv

Saved Model successfully.  
P8 - XGB Classifier: NSAP\_ML was successfully saved to predicting eligibility for NSAP program.  
[View in project](#)

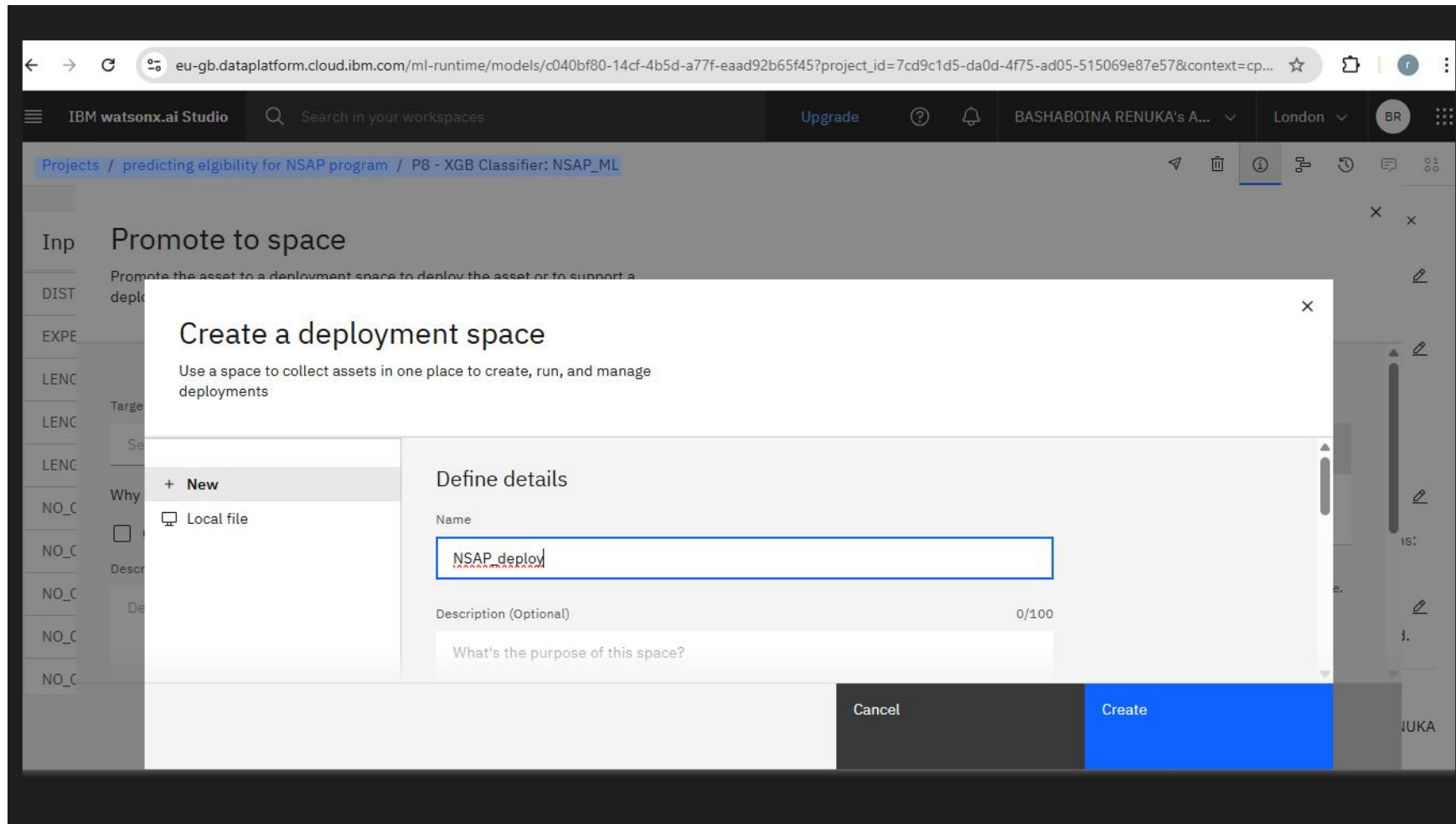
Pipeline leaderboard

	Rank ↑	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time	
★	1	<a href="#">Pipeline 8</a>	XGB Classifier		0.924	HPO-1 FE HPO-2	00:02:13	<a href="#">Save as</a>
	2	Pipeline 7	XGB Classifier		0.924	HPO-1 FE	00:01:28	
	3	Pipeline 6	XGB Classifier		0.918	HPO-1	00:00:29	
	4	Pipeline 5	XGB Classifier		0.918	None	00:00:03	

# 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





eu-gb.dataplatfom.cloud.ibm.com/ml-runtime/models/3efb7b91-31fb-4ce6-a3cf-e7bc7c8a5988/deployments?space\_id=4c25b102-0307-44da-ad69-7222ba4e706b...

IBM watsonx.ai Studio Search in your workspaces Upgrade ? 1 BASHABOINA RENUKA's A... London BR

Deployment spaces / NSAP\_deploy1 / P8 - XGB Classifier: NSAP\_ML

Deployments Model details

Search New deployment

Name	Type	Status	Tags	Last modified
NSAP_deploy1	Online	Deployed		1 minute ago BASHABOINA RENUKA (You)

Items per page: 20 1-1 of 1 items 1 of 1 pages

### About this asset

Name  
P8 - XGB Classifier: NSAP\_ML

Description  
No description provided.

Asset Details  
Type: wml-hybrid\_0.1  
Model ID: 3efb7b91-31fb-4c...  
Software specification: hybrid\_0.1  
Hybrid pipeline software specifications: autoai-kb\_rt24.1-py3.11

Tags  
Add tags to make assets easier to find.

Source asset details

---

# 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.

## NSAP\_deplyment ✓ Deployed Online

API reference

Test

Enter input data

Text

JSON

Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

:

Clear all ×

	STATE_NAME (other)	DISTRICT_NAME (other)	NO_OF_ROAD_WORK_SANCTIONED (double)	LENGTH_OF_ROAD_WORK_SANCTIONED (double)	NO_OF_BRIDGES
1	Start typing or drag and drop a CSV file...				
2					

0 rows, 14 columns

Predict

eu-gb.dataplatform.cloud.ibm.com/ml-runtime/deployments/c78ea561-10f8-44b9-8f69-1406e3626e97/test?space\_id=4c25b102-0307-44da-ad69-7222ba4e706b&c...


IBM watsonx.ai Studio Search in your workspaces Upgrade ? 1 BASHABOINA RENUKA's A... London BR

Deployment spaces / NSAP\_deploy1 / P8 - XGB Classifier: NSAP\_ML /

### NSAP Prediction results

Prediction type  
**Multiclass classification**

Prediction percentage



2 records

Display format for prediction results  
☒ Table view ☐ JSON view ☐ Show input data ⓘ

	Prediction	Confidence
1	RCPLWEA	94%
2	PMGSY-II	99%
3		
4		
5		
6		
7		
8		

Download JSON file

# 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

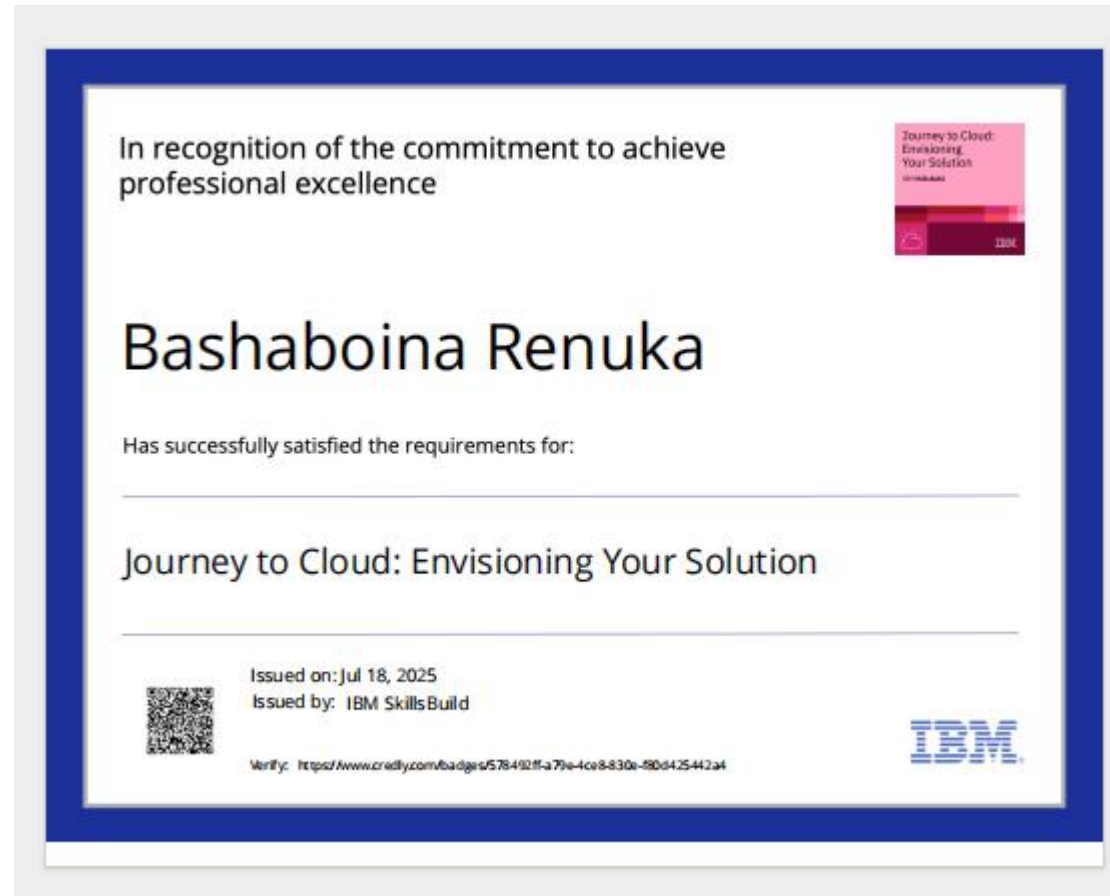
- AI-KOSH Dataset – PMGSY Project Data
- Government of India, National AI Portal
- URL: <https://aikosh.indiaai.gov.in>
- Scikit-learn: Machine Learning in Python
- Pedregosa et al., Journal of Machine Learning Research, 2011
- URL: <https://scikit-learn.org/>
- A key library used for classification 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](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

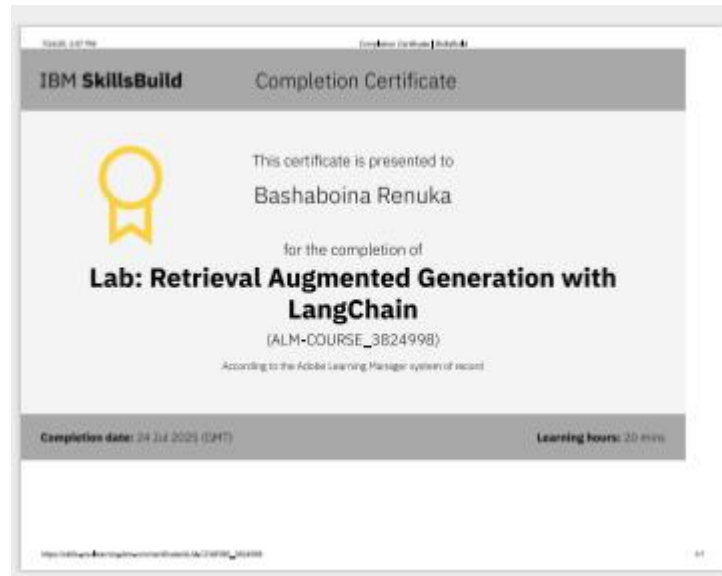




# IBM CERTIFICATIONS



# IBM CERTIFICATIONS





**THANK YOU**