# CAPSTONE PROJECT

## INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS USING MACHINE LEARNING

**Presented By:**
**1. Pujan C- Government Engineering College Rajkot – AI&DS**

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

Currently, the Pradhan Mantri Gram Sadak Yojana (PMGSY) operates under various sub-schemes like PMGSY-I, PMGSY-II, and RCPLWEA to improve rural infrastructure. Each project must be categorized under the appropriate scheme based on factors such as cost, location, and objectives. However, this classification is mostly done manually by officials, which is time-consuming, inconsistent, and prone to human error. As the number of infrastructure projects increases across districts and states, the burden of accurate classification grows, affecting planning and reporting accuracy. The crucial part is the intelligent classification of each project into its correct sub-scheme to streamline decision-making, resource allocation, and governance transparency.

# PROPOSED SOLUTION

- The proposed system aims to address the challenge of classifying rural infrastructure projects into their appropriate PMGSY sub-schemes using a machine learning-based approach. This will eliminate the manual process, reduce errors, and support efficient decision-making. The solution will consist of the following components:
- Data Collection:
  - Collect project-related data such as project name, cost, length, district, state, and funding details.
  - Use publicly available government datasets (e.g., AI Kosh PMGSY dataset).
- Data Preprocessing:
  - Clean and preprocess data to handle missing values, inconsistent formats, and outliers.
  - Perform feature engineering to extract relevant input features such as estimated cost range, geographical zone, or project duration.
- Machine Learning Algorithm:
  - Implement a multi-class classification algorithm such as Random Forest, Decision Tree, or IBM AutoAI's best-fit model.
  - Train the model using labeled data to predict whether a project belongs to PMGSY-I, PMGSY-II, or RCPLWEA.
- Deployment:
  - Deploy the trained model using IBM Watson Machine Learning as an API.
  - Create a simple web or desktop interface for users to upload project details and receive classification output.
- Evaluation:
  - Assess model performance using metrics such as Accuracy, Precision, Recall, and Confusion Matrix.
  - ine-tune based on model feedback to improve reliability across diverse datasets.
  - Result:

# SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the PMGSY scheme classification system. The goal is to design an intelligent and automated model that can classify projects based on their features with high accuracy and scalability.

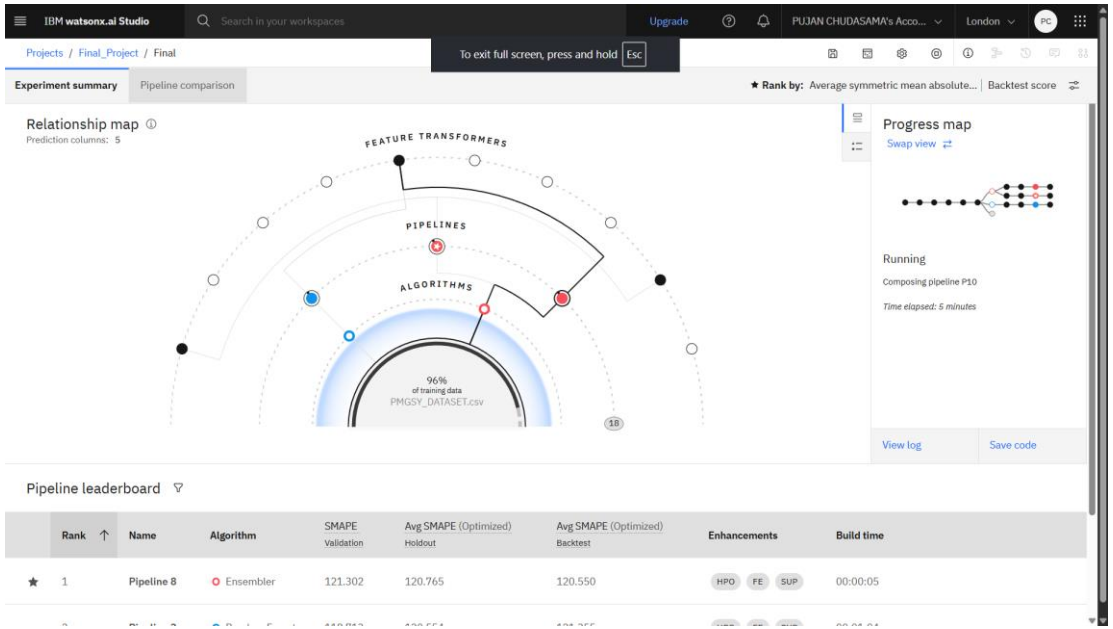
**System requirements :**

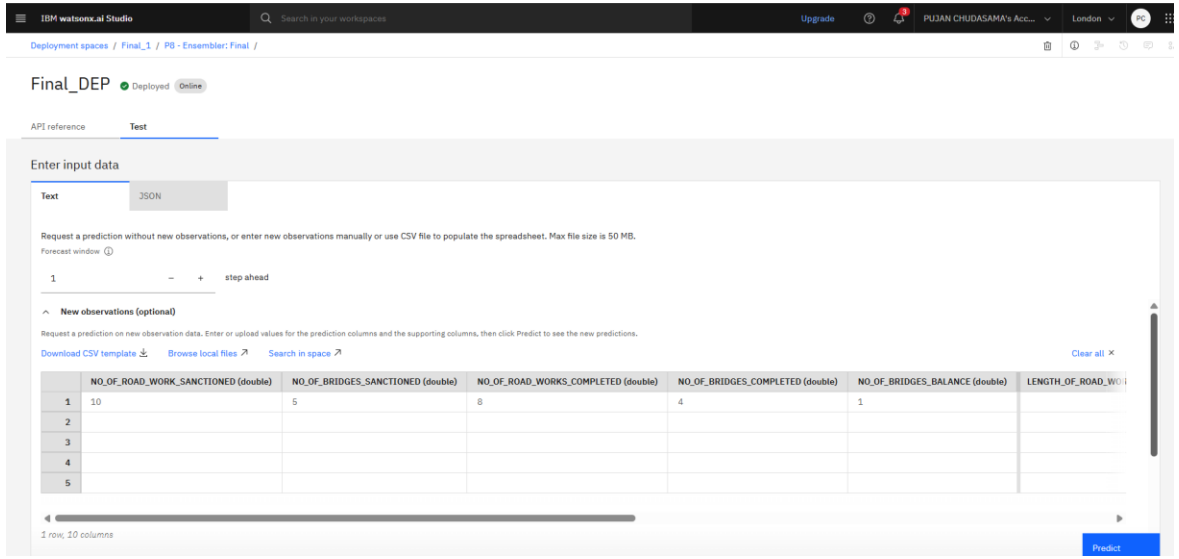
- A computing environment with internet access and a browser (for IBM Cloud platform).
- IBM Cloud Lite account with access to Watson Studio and AutoAI.
- Moderate processing capabilities for data handling and model training (e.g., 4 GB RAM minimum)
- Access to the PMGSY dataset from AI Kosh or data.gov.in.

# ALGORITHM & DEPLOYMENT

- In the Algorithm section, describe the machine learning algorithm chosen for predicting bike counts. Here's an example structure for this section:
- **Algorithm Selection:**
  - We used the **Random Forest Classifier** for its accuracy and ability to handle structured data. It reduces overfitting and works well for multi-class classification problems like this one.
- **Data Input:**
  - Number & length of sanctioned/completed road works
  - Project cost and expenditure.
  - State, district, and bridge counts.
- **Training Process:**
  - The model was trained using an 80/20 train-test split with cross-validation. Missing values were handled, and categorical data was encoded. Grid search was used for tuning hyperparameters.
- **Prediction Process:**
  - The trained model predicts the correct **PMGSY scheme** based on project details. It can be deployed via **IBM Watson Machine Learning** and used through a web-based interface for real-time classification.

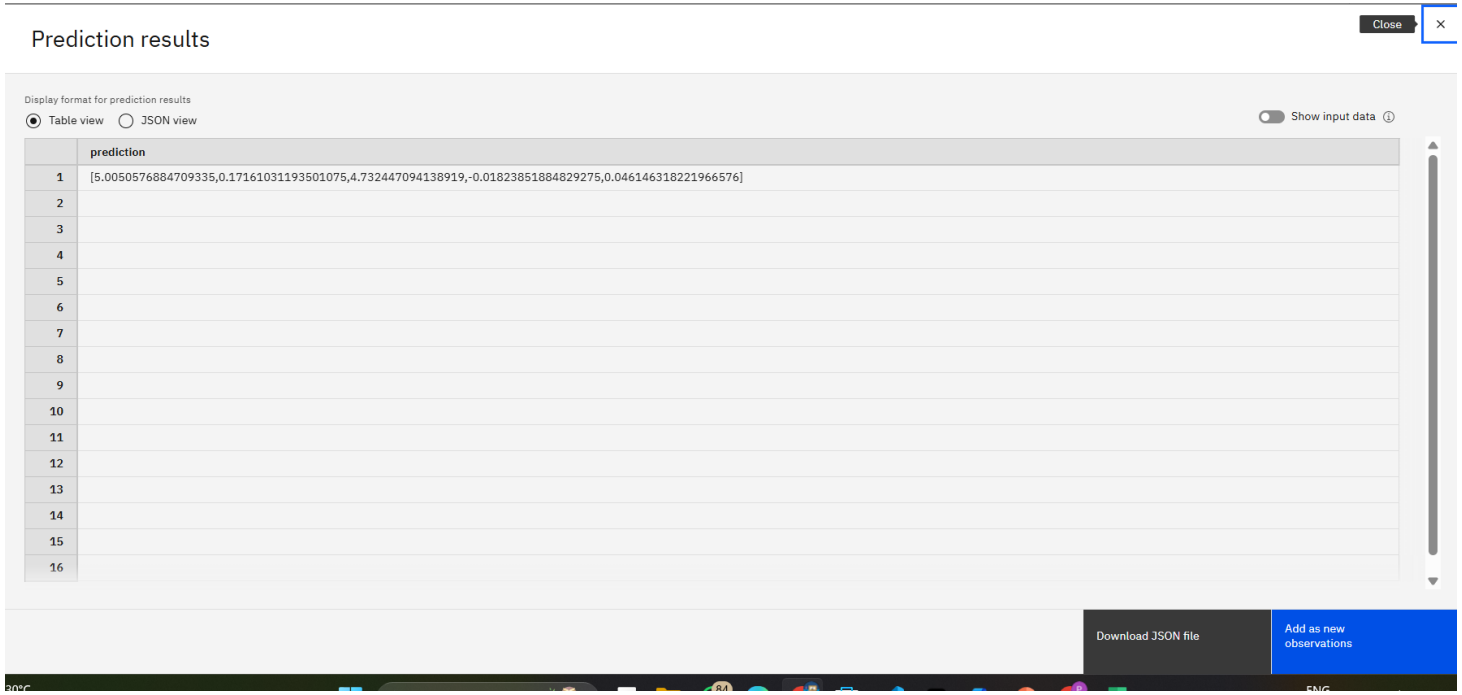
# RESULT




edunet foundation

# RESULT


Prediction results
Close
Display format for prediction results
Table view
JSON view
Show input data
prediction
1
[5.0050576884709335,0.17161031193501075,4.732447094138919,-0.01823851884829275,0.046146318221966576]
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Download JSON file
Add as new observations

# CONCLUSION

- The proposed machine learning model effectively classifies rural infrastructure projects into their respective PMGSY schemes based on physical and financial attributes. This automation improves accuracy, reduces manual effort, and supports better decision-making for rural development planning.
- During implementation, challenges included handling missing cost values and encoding categorical data. These were resolved through data preprocessing and feature engineering.
- The solution proves to be a reliable tool for accelerating project classification at scale. Accurate classification is essential for transparent reporting, efficient budget allocation, and ensuring accountability in large-scale government infrastructure programs.

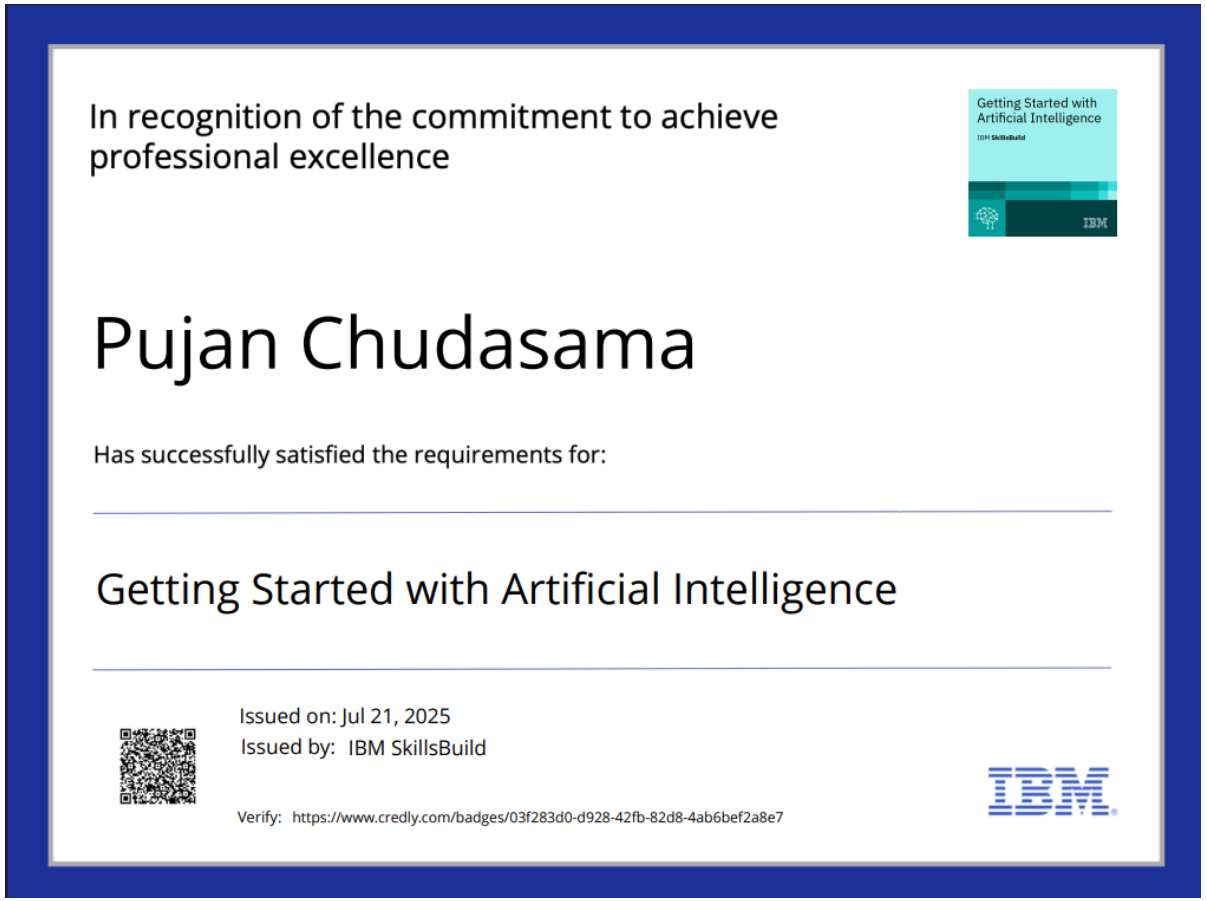
# FUTURE SCOPE

- The system can be enhanced by incorporating additional data sources such as geospatial data, satellite imagery, or real-time project updates from government portals. These inputs can improve model context and accuracy.
- Algorithm performance can be further optimized using advanced techniques such as ensemble stacking or deep learning models tailored to structured data.
- The system can also be scaled to support **all states and union territories**, enabling broader infrastructure planning coverage. Additionally, integration with **edge computing platforms** can help local authorities make faster, offline predictions in rural areas.
- Future upgrades may also include automated data ingestion pipelines and interactive dashboards for policymakers and field engineers.

# REFERENCES


- **AI Kosh PMGSY Dataset**
  aikosh.indiaai.gov.in
- **Scikit-learn Library**
  scikit-learn.org
- **IBM Watson Studio Documentation**
  dataplatform.cloud.ibm.com
- **Random Forests – Breiman (2001)**
  Machine Learning Journal
- **Machine Learning Preprocessing Guide**
  machinelearningmastery.com

# IBM CERTIFICATIONS



# IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM SkillsBuild

IBM

Pujan Chudasama

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

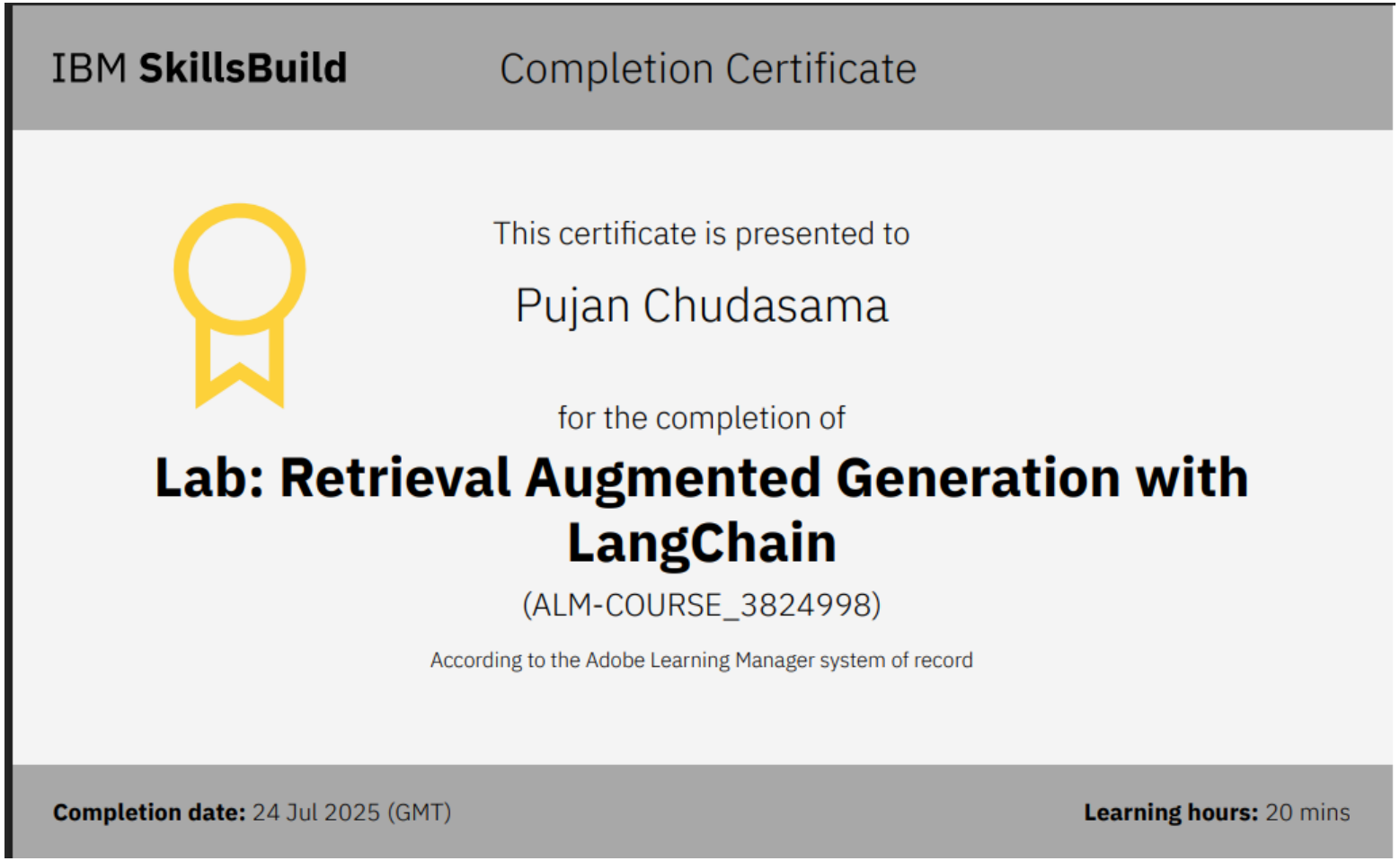
Issued on: Jul 21, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/feaefff4-f97b-4356-ae0c-c09a4ce0265f

IBM

# IBM CERTIFICATIONS



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Pujan Chudasama

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