CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:
1. Ritam Naskar- Jadavpur University-Information Technology

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

Example: The Pradhan Mantri Gram Sadak Yojana (PMGSY) is a Government of India initiative aimed at providing all-weather road connectivity to unconnected rural habitations. Over the years, the program has evolved into multiple schemes, including PMGSY-I, PMGSY-II, and RCPLWEA, each with different objectives, funding patterns, and technical parameters.With thousands of projects being implemented across various regions, manual classification of these projects into their respective PMGSY schemes is inefficient, time-consuming, and error-prone. This creates challenges in budget allocation, impact evaluation, and progress monitoring.There is a critical need for a system that can automatically categorize a project based on its physical and financial features into the correct PMGSY scheme, enabling better decision-making, analytics, and governance.

PROPOSED SOLUTION

- The proposed system is an AI-based machine learning model capable of classifying rural infrastructure projects (roads or bridges) into their respective PMGSY schemes using their physical and financial attributes.
- Key features:
- Uses structured data provided by government sources like AI Kosh.
- Automatically identifies the project category (PMGSY-I, PMGSY-II, RCPLWEA).
- Built using supervised machine learning techniques.Deployed using IBM Cloud Lite services, specifically Watson Studio, Cloud Object Storage, and Jupyter Notebooks.
- Offers an intelligent, scalable, and reliable alternative to manual classification.

SYSTEM APPROACH

a) Dataset AcquisitionSource: Al Kosh – Pradhan Mantri Gram Sadak Yojna (PMGSY) Dataset.Contains information on financial outlays, physical parameters, state info, project status, and scheme types.

b) Tools and Technology UsedIBM Watson Studio (Jupyter Notebook for model development).IBM Cloud Object Storage (for dataset storage).Python Libraries: pandas, sklearn, matplotlib, seaborn.

c) Data PreprocessingRemoval of missing values.Label encoding of categorical variables.Feature-target split: Features = all attributes; Target = PMGSY_SCHEME.

d) Model Selection and TrainingAlgorithm: Random Forest Classifier (due to robustness and interpretability).Dataset split into training (80%) and test (20%) subsets.Model trained on the training set and evaluated on the test set.

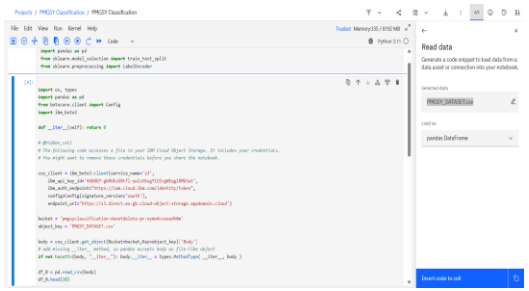
e) Evaluation MetricsAccuracy ScoreClassification Report (Precision, Recall, F1-Score)Confusion Matrix visualization

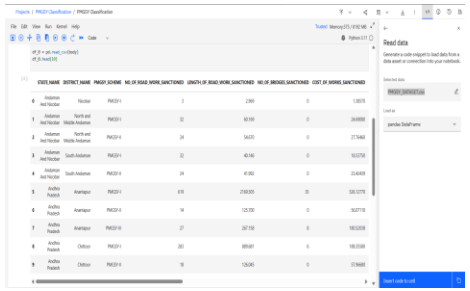
f) DeploymentIBM Cloud-based Jupyter Notebook.Model code saved and exported as part of project deliverables.

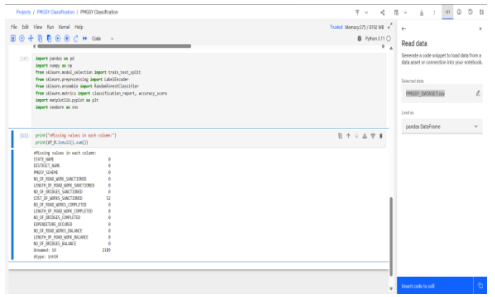
ALGORITHM & DEPLOYMENT

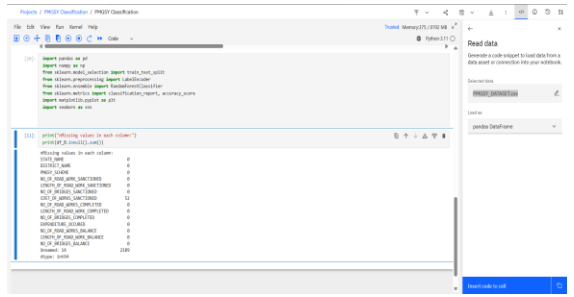
- The solution follows a supervised machine learning classification approach. The goal is to predict the correct PMGSY scheme type for a given rural infrastructure project based on its physical and financial features.
- Step-by-Step Algorithm Description:
- 1. Data CollectionCollect the PMGSY dataset from AI Kosh, which contains various attributes of rural road/bridge construction projects such as state, district, cost, project length, date, etc.
- 2. Data PreprocessingClean the dataset by removing null or irrelevant values.Convert categorical variables (like state, district, project type) into numerical form using encoding techniques such as Label Encoding or One-Hot Encoding.Normalize or scale numerical values if required to bring them to a common range.
- 3. Feature SelectionIdentify and select key physical and financial features that influence the classification of the project into its correct PMGSY scheme (PMGSY-I, PMGSY-II, RCPLWEA, etc.).
- 4. Splitting the DatasetSplit the dataset into training and testing subsets, typically using an 80:20 ratio. This helps in training the model and evaluating its performance
- .5. Model Selection and TrainingChoose a suitable classification algorithm such as Random Forest Classifier, which works well for structured tabular data.Train the model on the training data by feeding in the input features and their corresponding scheme labels
- .6. Model EvaluationTest the trained model using the test dataset to evaluate its prediction accuracy.Use performance metrics such as accuracy score, confusion matrix, precision, recall, and F1-score to assess the model's effectiveness.
- 7. Model SavingSave the trained model to a file so it can be reused or deployed later without retraining. DEPLOYMENT STRATEGY (WRITTEN EXPLANATION)The deployment of this ML project is done using IBM Cloud Lite services, particularly IBM Watson Studio and IBM Cloud Object Storage.Step-by-Step Deployment Process:
- 1. Set up IBM Cloud EnvironmentCreate a free IBM Cloud Lite account.Launch IBM Watson Studio and create a new project.Attach a Cloud Object Storage instance to the project for storing data assets and models
- .2. Upload DatasetAdd the PMGSY dataset to the project under "Assets" as a Data Asset.This will allow direct access from your Jupyter notebook within the IBM Watson Studio.
- 3. Create and Run a Jupyter NotebookCreate a new notebook inside Watson Studio using Python.Import required libraries, load the dataset from Cloud Object Storage, preprocess it, train the model, and evaluate its performance.
- 4. Save the ModelAfter successful training and evaluation, serialize the model using joblib or pickle and save it as a .pkl file to Cloud Object Storage or local environment for future use.
- 5. Build a Prediction Service (Optional)Develop a REST API using Flask (locally or on IBM Cloud) to accept input features and return the predicted PMGSY scheme.This allows other applications or stakeholders to use the trained model for real-time predictions.
- 6. Deploy the Model on IBM Cloud (Optional Advanced Step)If needed, containerize the model using Docker.Deploy it as a microservice using IBM Cloud Code Engine, Cloud Foundry, or IBM Cloud Functions.Provide a secure API endpoint that accepts inputs and returns predictions.

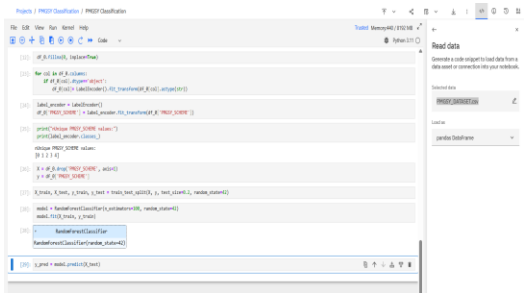
RESULT

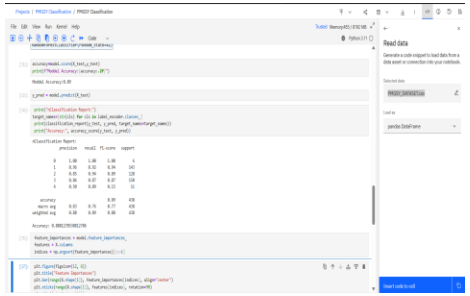


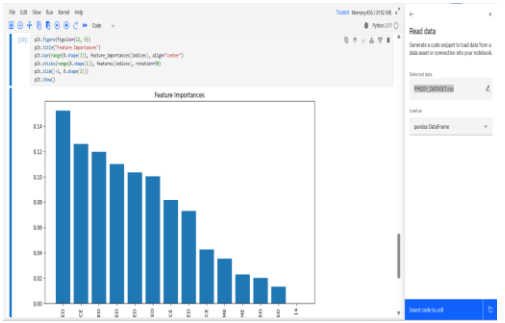


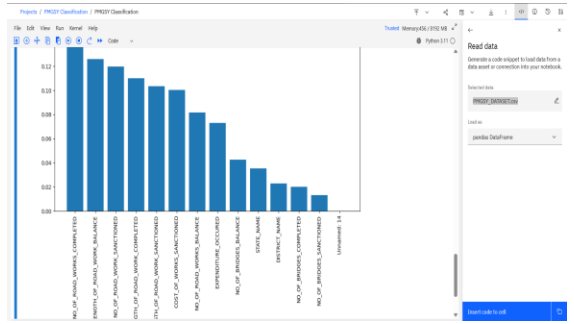












edunet
foundation

CONCLUSION

- This project successfully demonstrates how artificial intelligence can streamline the classification of rural infrastructure projects under PMGSY. The model built using machine learning algorithms accurately classifies projects into respective schemes based on input features.By automating the classification process, the system reduces human error, increases scalability, and enhances decision-making for government bodies and policy analysts involved in rural development.

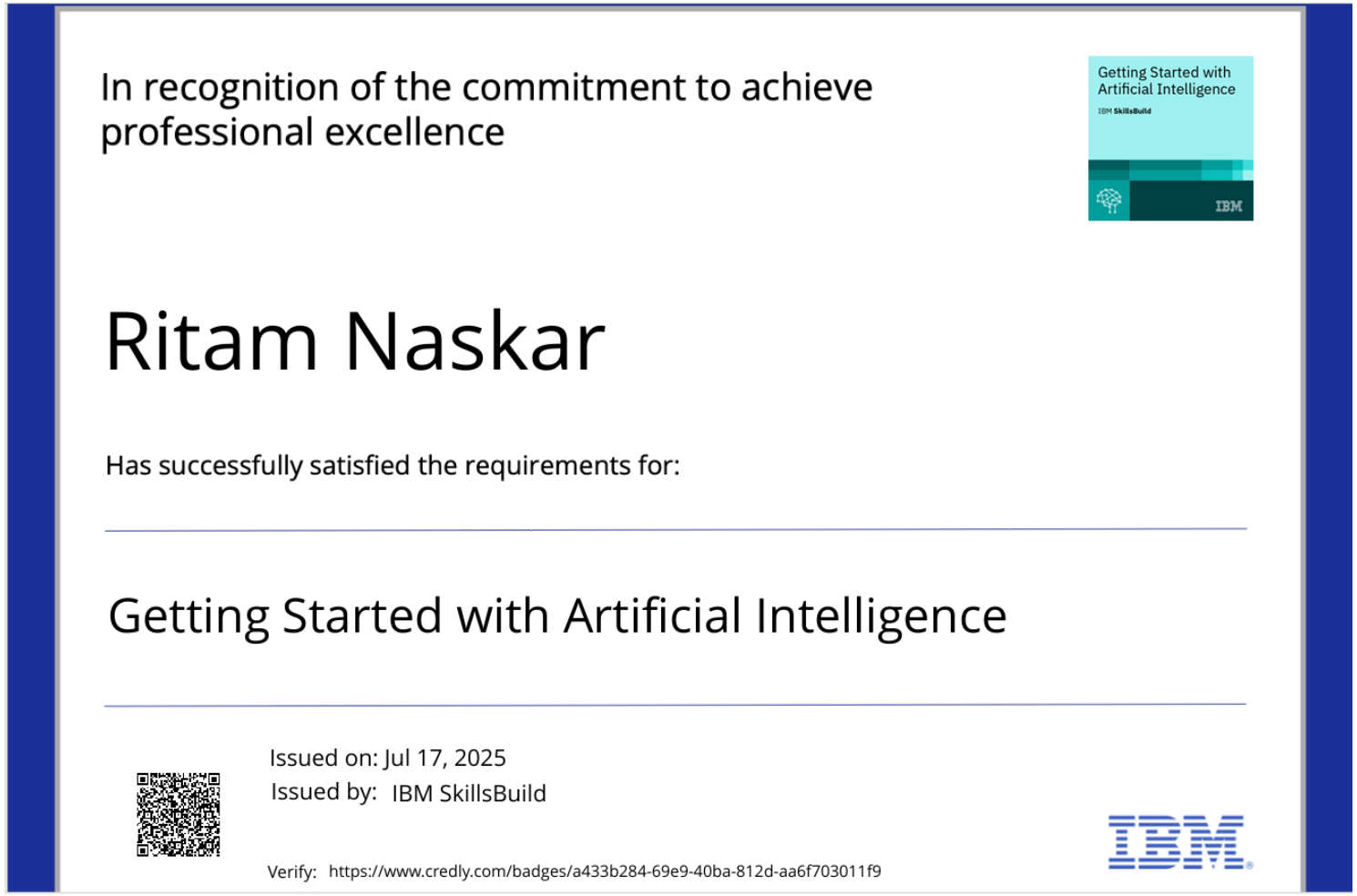
FUTURE SCOPE

- Integration with real-time dashboards for district/block-level monitoring.Deployment as a web or mobile app using IBM Cloud Functions or Flask.Incorporation of geospatial data (satellite maps, GPS) to improve accuracy.Multi-language support for regional data entry operators.Extension to other schemes (e.g., PMAY, Jal Jeevan Mission) using similar frameworks.

REFERENCES

- 1. AI Kosh Dataset – PMGSY:
- https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html
- 2. Pradhan Mantri Gram Sadak Yojana Official Site:
- https://pmgsy.nic.in
- 3. IBM Watson Studio:
- https://dataplatform.cloud.ibm.com
- 4. Scikit-learn Documentation:
- https://scikit-learn.org/stable/
- 5. IBM Cloud Object Storage Documentation:
- https://cloud.ibm.com/docs/cloud-object-storage

IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild

Ritam Naskar

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

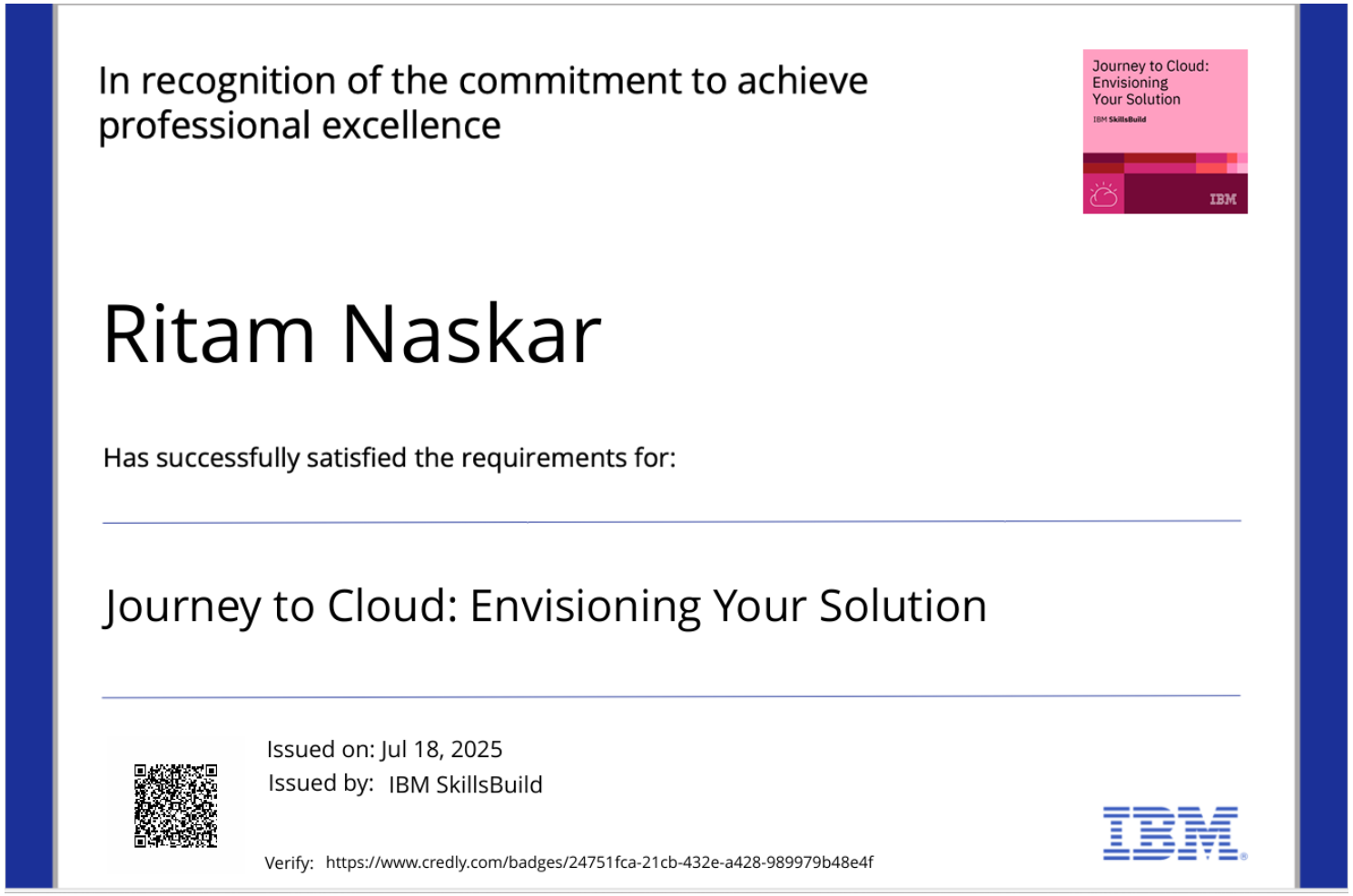
Issued on: Jul 17, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/a433b284-69e9-40ba-812d-aa6f703011f9

IBM

edunet foundation

IBM CERTIFICATIONS



In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild

Ritam Naskar

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 18, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/24751fca-21cb-432e-a428-989979b48e4f

IBM

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Ritam Naskar

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT) **Learning hours:** 20 mins

THANK YOU

edunet foundation