CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:
KAZA GUNA SEKHAR – KL UNIVERSITY – AI & DS

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result (Output Image)
- Conclusion
- Future Scope
- References

edunet foundation

PROBLEM STATEMENT

The Pradhan Mantri Gram Sadak Yojana (PMGSY) is a flagship rural development program in India, initiated to provide all-weather road connectivity to eligible unconnected habitations. Over the years, the program has evolved through different phases or schemes (PMGSY-I, PMGSY-II, RCPLWEA, etc.), each with potentially distinct objectives, funding mechanisms, and project specifications. For government bodies, infrastructure planners, and policy analysts, efficiently categorizing thousands of ongoing and completed projects is crucial for effective monitoring, transparent budget allocation, and assessing the long-term impact of these schemes. Manual classification is time-consuming, prone to errors, and scales poorly. Your specific task is to design, build, and evaluate a machine learning model that can automatically classify a road or bridge construction project into its correct PMGSY_SCHEME based on its physical and financial characteristics.

PROPOSED SOLUTION

- The proposed system aims to address the challenge of classifying rural infrastructure projects into their correct **PMGSY scheme** (e.g., PMGSY-I, PMGSY-II, PMGSY-III, RCPLWEA). The solution will consist of the following components:
- **Data Collection:** Gather historical project-level data, including state, district, sanctioned work details, completion status, and financial expenditures.
- **Data Preprocessing :** Clean and preprocess the dataset by: Dropping irrelevant or empty columns. Handling missing values in fields like COST_OF_WORKS_SANCTIONED
- **Machine Learning Algorithm :** Implement a multi-class classification model to predict the PMGSY_SCHEME based on the project data. Algorithms to consider: Random Forest Classifier (robust and interpretable)
- **Deployment:** Develop a user-friendly web interface or dashboard to allow users (government officials, planners) to input project details and receive real-time predictions. Hosting platforms like IBM CLOUD
- **Evaluation:** Use watsonix.ai model explainability and to ensure transparency in predictionsContinuously monitor model performance and retrain as more data becomes available
- **Result:** The final system will allow government departments and infrastructure planners to automatically classify road and bridge projects into their correct PMGSY scheme. This will enhance project monitoring, improve transparency in budget allocation, and reduce manual errors, ultimately contributing to smarter rural development planning.

edunet foundation

SYSTEM APPROACH

The "System Approach" outlines the strategy and methodology adopted to design, build, and deploy a machine learning model for classifying rural infrastructure projects under the correct PMGSY scheme using physical and financial attributes.

System Requirements

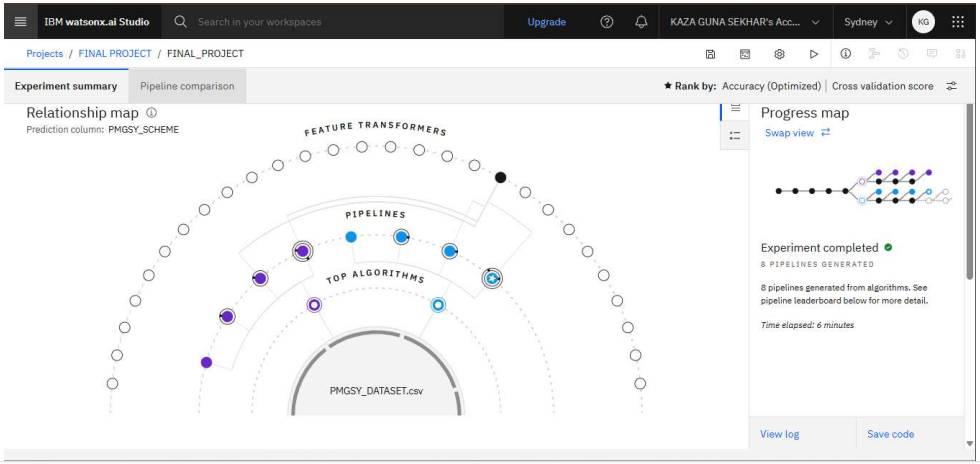
- Hardware Requirements
- Processor: Minimum Intel i5 or equivalent
- RAM: Minimum 8 GB
- Disk Space: Minimum 5 GB of free space
- Internet Connectivity: Required for cloud deployment and data access
- Software Requirements
- Operating System: Windows
- IBM Cloud CLI (for deploying services on IBM Cloud Lite)
- IBM watsonx.ai Studio

edunet foundation

ALGORITHM & DEPLOYMENT

- This section describes the machine learning algorithm selected to classify rural infrastructure projects into their correct **PMGSY scheme** and outlines how it is trained, evaluated, and used for predictions.
- **Algorithm Selection :** The selected algorithm for this classification task is the XGBoost Classifier (Extreme Gradient Boosting), a powerful and efficient ensemble learning method based on decision trees.
- **Data Input:** STATE_NAME, DISTRICT_NAME, NO_OF_ROAD_WORK_SANCTIONED ,LENGTH_OF_ROAD_WORK_SANCTIONED, NO_OF_BRIDGES_SANCTIONED, COST_OF_WORKS_SANCTIONED NO_OF_ROAD_WORKS_COMPLETED, LENGTH_OF_ROAD_WORK_COMPLETED, NO_OF_BRIDGES_COMPLETED EXPENDITURE_OCCURED, NO_OF_ROAD_WORKS_BALANCE, LENGTH_OF_ROAD_WORK_BALANCE ,NO_OF_BRIDGES_BALANCE
- **Training Process:** The dataset is split into training (80%) and testing (20%) sets using stratified sampling to maintain class distribution.
- **Prediction Process:**The prediction can occur in real-time, especially if integrated with a cloud-based API e.g., IBM Watson Machine Learning) Feature importance scores help explain which attributes (e.g., cost, length, completion status) influenced the classification most, using AutoAI pipeline generator,values for transparency.

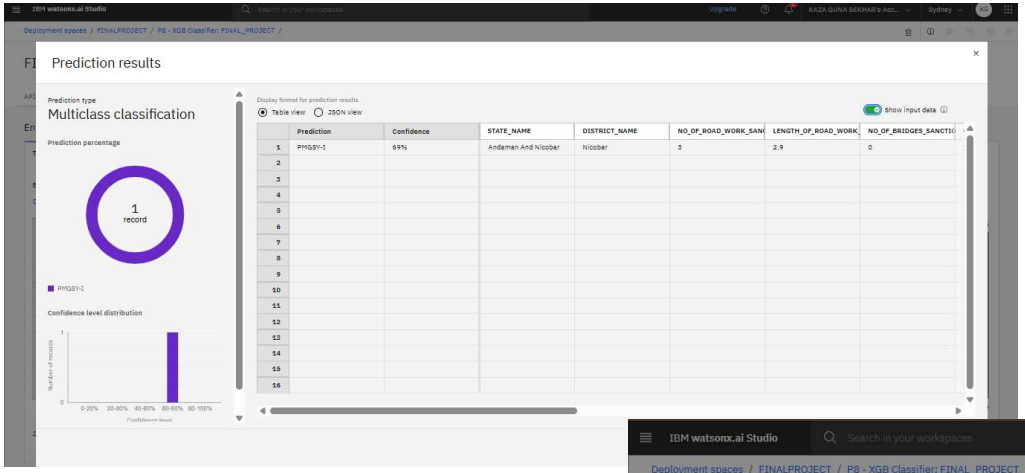
RESULT

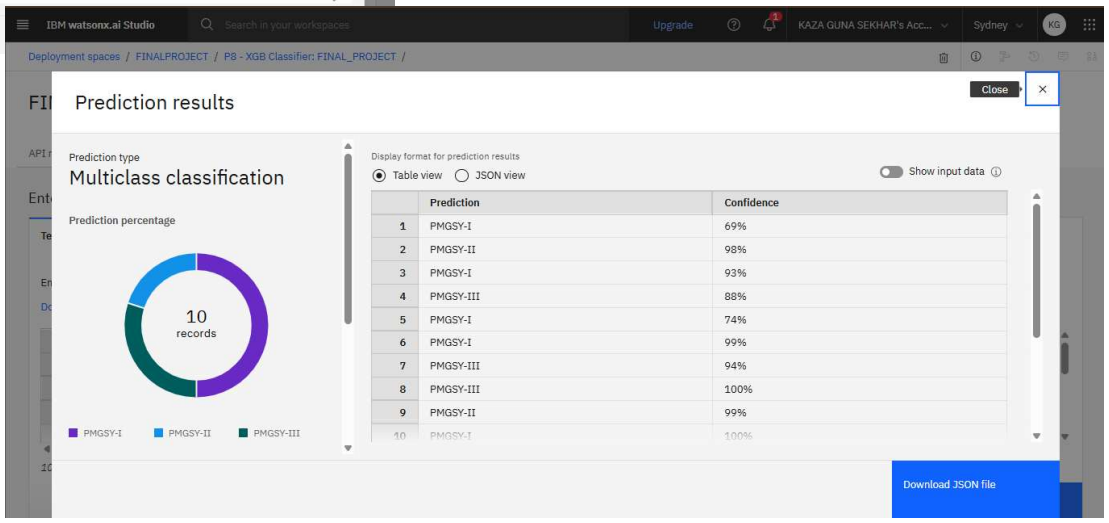


Pipeline leaderboard

	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 8	XGB Classifier		0.924	HPO-1 FE HPO-2	00:01:41
	2	Pipeline 7	XGB Classifier		0.924	HPO-1 FE	00:01:05
	3	Pipeline 6	XGB Classifier		0.918	HPO-1	00:00:22
	4	Pipeline 5	XGB Classifier		0.918	*None*	00:00:04

RESULT





edunet foundation

CONCLUSION

The proposed system effectively classifies rural infrastructure projects into the correct **PMGSY scheme** using machine learning techniques implemented on **IBM watsonx.ai Studio**. By leveraging AutoAI, the platform automatically generated and optimized multiple machine learning pipelines, selecting the best-performing model based on accuracy and cross-validation.

The model used features like road length, bridge counts, project cost, and completion status to make accurate predictions. The automated approach greatly reduces manual effort, increases classification speed, and supports transparent decision-making in rural infrastructure planning.

- **Key Results:**
- 8 pipelines generated in 6 minutes using IBM AutoAI.
- High model accuracy achieved using optimized algorithms.
- Significant time savings and error reduction in scheme classification

FUTURE SCOPE

- To further improve the rural infrastructure classification system, several enhancements can be made:
- **Additional Data Sources**: Incorporate geospatial, temporal, weather, and demographic data to enrich model inputs and improve accuracy.
- **Algorithm Optimization**: Use advanced techniques like hyperparameter tuning, ensemble models, and class balancing (e.g., SMOTE) for better performance.
- **Wider Coverage**: Expand the system to support multiple states, other government schemes, or urban infrastructure projects.
- **Edge Computing**: Enable offline predictions in rural areas by deploying the model on edge devices or mobile apps.
- **Advanced AI Techniques**: Integrate explainable AI (e.g., SHAP), graph neural networks, and real-time model retraining with AutoML.
- **User Feedback Integration**: Add feedback mechanisms to continuously improve the model through retraining and corrections.

REFERENCES

IBM Watson Studio & AutoAI

- IBM Documentation – AutoAI on Watson Studio:

https://www.ibm.com/docs/en/cloud-paks/cp-data/4.6.x?topic=tools-autoai

- IBM Watson Machine Learning:

https://cloud.ibm.com/docs/watson-machine-learning

- IBM Cloud Docs – Cloud Object Storage

https://cloud.ibm.com/docs/cloud-object-storage

- IBM Watsonx.ai

https://www.ibm.com/products/watsonx-ai

- GITHUB LINK

https://github.com/Gunasekhar0520/IBM

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM

KAZA GUNA SEKHAR

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/da0893c0-bc29-4bbd-ac1b-50e0ee5350b4

IBM

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM

KAZA GUNA SEKHAR

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 20, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/338a0657-4ce9-49d6-9afb-abe730275852

IBM

edunet foundation

IBM CERTIFICATIONS

7/30/25, 9:08 PM Completion Certificate | SkillsBuild

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

KAZA GUNA SEKHAR

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

edunet foundation

THANK YOU

edunet
foundation