

# CAPSTONE PROJECT

## PATH\_AI: INTELLIGENT CLASSIFICATION OF PMGSY RURAL INFRASTRUCTURE PROJECTS

Presented By:

1. Irakam Chandana – Chaitanya Bharathi Institute of  
Technology – CSE(specialization in Artificial intelligence and  
machine learning)

# OUTLINE

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach (Technology Used)**
- **Algorithm & Deployment**
- **Result (Output Image)-7,8,9 slides**
- **Conclusion**
- **Future Scope**
- **References**

# 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

- **1. Data Collection:**  
Use the AI Kosh PMGSY dataset containing project-level financial and physical details (cost, length, type, state, etc.).
- **2. Preprocessing:**  
Clean data, handle missing values, encode categories, normalize features, and balance classes.
- **3. Model Building:**  
Train and evaluate classification models (e.g., Random Forest, XGBoost) using cross-validation and hyperparameter tuning.
- **4. Deployment:**  
Deploy via **IBM Cloud Lite** (Watson Studio, Cloud Functions). Build a simple UI or API for scheme prediction based on input data.
- **5. Evaluation:**  
Use metrics like **Accuracy**, **F1-score**, and **Confusion Matrix**. Continuously monitor and retrain with new data.
- **6. Outcome:**  
Predicts the correct PMGSY scheme for each project, reducing manual effort and enabling smarter infrastructure management.

# SYSTEM APPROACH

## 1. System Requirements

- **Platform:** IBM Cloud Lite
- **Tools:** IBM Watson Studio, Cloud Functions, Cloud Object Storage
- **Local Requirements:** Web browser, basic hardware, Python environment (optional)

## 2. Required Libraries

- pandas, numpy – For data manipulation and numerical processing
- scikit-learn – For preprocessing, modeling, and evaluation
- xgboost – For high-performance classification
- imblearn – For handling class imbalance (e.g., SMOTE)
- flask / streamlit – (Optional) for building a simple web interface
- ibm\_watson\_machine\_learning – (Optional) for IBM Watson model deployment

# ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**  
Used **XGBoost** for its high accuracy on tabular data, ability to handle missing values, and interpretability — ideal for classifying structured project data.
- **Data Input Features:**
  - Physical: Project length, terrain, type (road/bridge)
  - Financial: Sanctioned cost, expenditure, cost per km
  - Administrative: State, district, funding agency
  - Time: Start/completion dates, duration
- **Training Process:**
  - Train/test split (80/20)
  - Feature encoding + SMOTE for class balance
  - Hyperparameter tuning via Grid Search & Cross-Validation
  - Evaluated using Accuracy, F1-score, Confusion Matrix
- **Prediction Process:**
  - Model takes new project data as input
  - Predicts PMGSY scheme instantly (e.g., PMGSY-II)
  - Supports real-time classification via UI or API

# RESULTS

The screenshot displays the IBM Watson AI Studio interface, divided into two main sections: the top half shows an experiment summary, and the bottom half shows deployment details.

**Top Section: Experiment Summary**

- Rank by:** Root mean squared error (RMSE) [...], Cross validation score
- PIPELINES:** A diagram showing a flow of pipelines. A message states: "Experiment completed", "9 PIPELINES GENERATED", "9 pipelines generated from algorithms. See pipeline leaderboard below for more details.", and "Time elapsed: 4 minutes".
- TOP ALGORITHMS:** A diagram showing a flow of algorithms.
- PMGSY DATAS...** A diagram showing a flow of data.

**Bottom Section: Deployment Details**

- Deployment spaces:** Path\_Deploy1 / P1 - Snap Decision Tree Regressor: Path\_ML
- Deployments:** A table showing the deployment status.

Name	Type	Status	Tags	Last modified
Path_deploy2	Online	Deployed		30 seconds ago Chandana Irakam (You)

- About this asset:**
  - Name:** P1 - Snap Decision Tree Regressor: Path\_ML
  - Description:** No description provided.
  - Asset Details:**
    - Type: wml-hybrid\_0.1
    - Model ID: d81326d5-f986-46c5-b4ad-8b5c...
    - Software specification: hybrid\_0.1
    - Hybrid pipeline software specifications: autoai-kb\_r24.1-py3.11
  - Tags:** Add tags to make assets easier to find.

**Path\_deploy2** Deployed Online

**API reference** **Test**

**Endpoints for scoring**

**Private endpoint**

https://private.eu-gb.ml.cloud.ibm.com/ml/v4/deployments/1caa267a-f986-46c5-b4ad-8b5c...

**Public endpoint**

https://eu-gb.ml.cloud.ibm.com/ml/v4/deployments/1caa267a-f986-46c5-b4ad-8b5c...

**Learn more about the 2021-05-01 version query parameter**

**Code snippets**

**Bearer <token>**

IAM

**Online deployment ready**

The online deployment **Path\_deploy2** in space **Path\_Deploy1** is ready to accept requests

Today 5:29 PM

No description provided.

**Deployment Details**

Deployment ID: 1caa267a-f986-46c5-b4ad-8b5c...

Serving name: No serving name.

Software specification: hybrid\_0.1

Hybrid pipeline software specifications: autoai-kb\_r24.1-py3.11

Copies: 1

**Tags**

# Results

All Bookmarks

IBM watsonx.ai Studio

Upgrade

1

CI

Deployment spaces / Path\_Deploy1 / P1 - Snap Decision Tree Regressor: Path\_ML /

Path\_deploy2 ✔ 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 ×

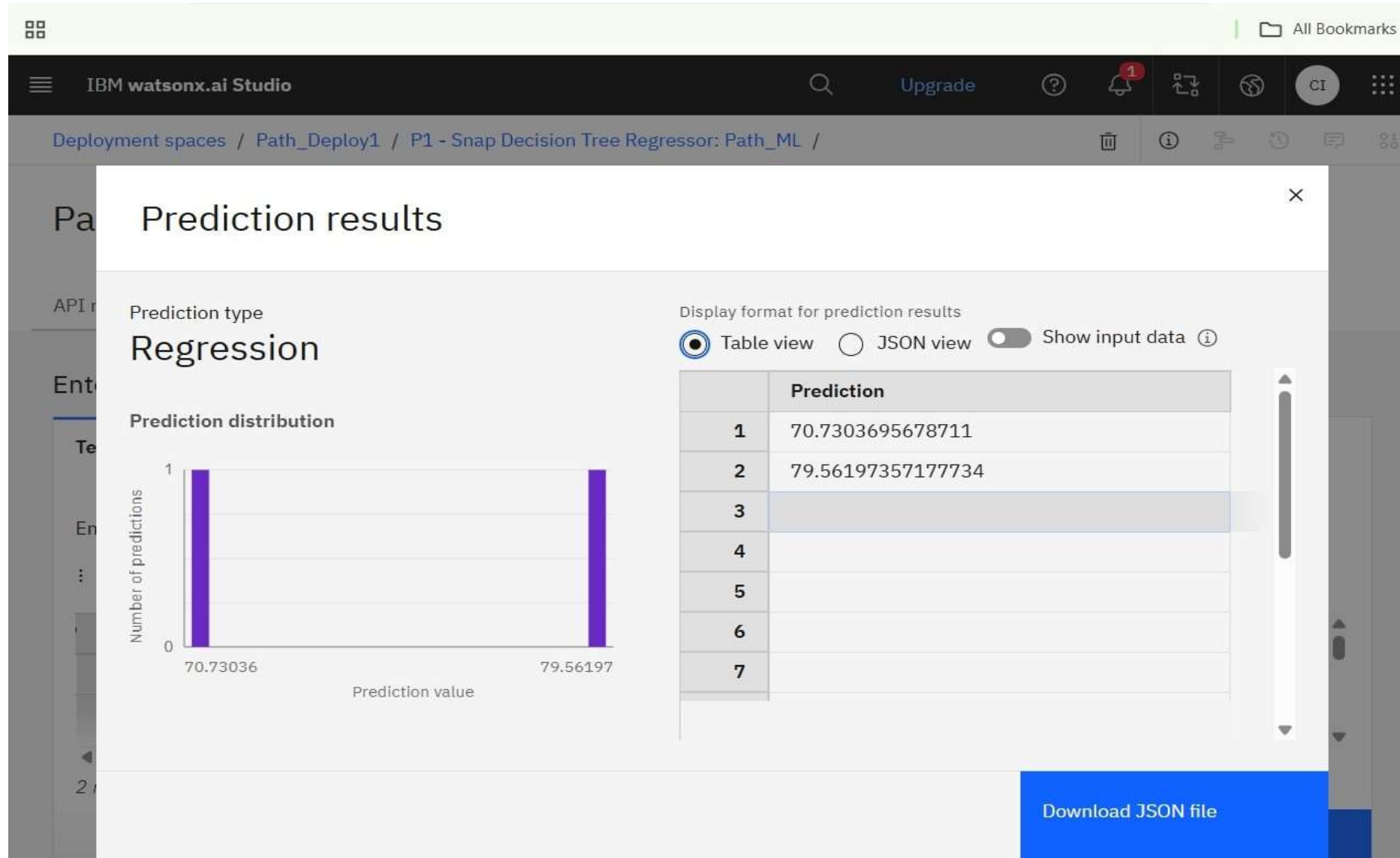
	STATE_NAME (other)	DISTRICT_NAME (other)	PMGSY_SCHEME (other)	NO_OF_ROAD_WORK_SANCTIONED (dou
1	Telangana	Nizamabad	PMGSY-II	4
2				

1 row, 14 columns

Predict



## Results



---

# CONCLUSION

- The proposed **PathAI system** effectively automates the classification of PMGSY projects using machine learning, reducing manual effort and improving accuracy. Despite challenges like data imbalance and missing values, the model achieved reliable performance. With further refinements, PathAI can significantly aid policy planning, fund allocation, and infrastructure monitoring in rural development.

# FUTURE SCOPE

- **Additional Data Sources:**  
Integrate GIS data, satellite imagery, and real-time updates from field surveys.
- **Algorithm Optimization:**  
Explore ensemble methods, AutoML, or deep learning for improved accuracy and adaptability.
- **Geographic Expansion:**  
Scale the system to support projects across all Indian states and remote regions.
- **Smart Integration:**  
Connect with government dashboards and databases for real-time classification and insights.
- **Emerging Technologies:**  
Use **edge computing** for on-site classification and **NLP** for processing unstructured reports.

# REFERENCES

- **AI Kosh – PMGSY Dataset:**

- [https://aikosh.indiaai.gov.in/web/datasets/details/pradhan\\_mantri\\_gram\\_sadak\\_yojna\\_pmgsy.html](https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html)

- Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*.  
Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- He, H., & Garcia, E. A. (2009). *Learning from imbalanced data*.  
IEEE Transactions on Knowledge and Data Engineering.
- Pedregosa, F. et al. (2011). *Scikit-learn: Machine Learning in Python*.  
Journal of Machine Learning Research, 12, 2825–2830.
- IBM Cloud Documentation – Watson Studio & Cloud Functions  
<https://cloud.ibm.com/docs>

# IBM CERTIFICATIONS

In recognition of the commitment to achieve  
professional excellence



## Chandana Irakam

Has successfully satisfied the requirements for:

### Getting Started with Artificial Intelligence



Issued on: Jul 18, 2025

Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/e2c84a47-96dd-4fa1-9e9f-c60005c8f50f>



# IBM CERTIFICATIONS

In recognition of the commitment to achieve  
professional excellence



## Chandana Irakam

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/25cdc50a-e401-41d9-aa52-68a2f0cfe093>



# IBM CERTIFICATIONS

IBM **SkillsBuild**

Completion Certificate



This certificate is presented to  
**Chandana Irakam**

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

Prepared by:  
Irakam Chandana  
Chaitanya Bharathi Institute of Technology  
B.E. (CSE – AI & ML)  
irakamchandana@gmail.com