CAPSTONE PROJECT

PMGSY SCHEME CLASSIFIER

Presented By:
Aakash Singh
IET, Dr. Shakuntala Misra National Rehabilitation University Lucknow
Department of Electronics and Communication Engineering

edunet foundation

OUTLINE

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

PROBLEM STATEMENT

In India, the Pradhan Mantri Gram Sadak Yojana (PMGSY) plays a vital role in connecting rural communities by building all-weather roads and bridges. Over time, this program has introduced multiple phases, each with its own goals and funding methods. With thousands of projects underway or completed, keeping track of which project belongs to which phase has become challenging. Doing this manually takes a lot of time, is prone to mistakes, and doesn't scale well. To support better planning, budgeting, and evaluation, we need an intelligent automated system that can quickly and accurately classify these projects based on their characteristics. This would help decision-makers monitor progress and make informed choices to improve rural connectivity.

PROPOSED SOLUTION

- The proposed system aims to streamline and automate the classification of road and bridge construction projects into their respective PMGSY schemes. This solution leverages data science and machine learning to accurately categorize projects based on physical and financial parameters. The main components are as follows:
- Data Collection:
  - Gather historical data on completed and ongoing road and bridge projects, including their physical specifications, financial details, project status, and existing scheme labels.
  - Source additional relevant datasets, such as government records or the AI Kosh dataset, to ensure diversity and accuracy in the training data.
- Data Preprocessing:
  - Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.
  - Perform feature engineering to identify and construct the most impactful features for distinguishing between different PMGSY schemes.
- Machine Learning Algorithm:
  - Utilize IBM AutoAI to automate the selection, training, and optimization of machine learning classification models suited to this problem.
  - Let the model learn from financial, physical, and administrative features to predict the appropriate scheme for each project.
- Deployment:
  - Deploy the trained model on IBM Cloud Lite services, ensuring scalability and accessibility for end-users, planners, and policy makers.
  - Integrate the solution into an interface or API, allowing stakeholders to quickly classify new projects as data becomes available.
- Evaluation:
  - Assess model performance using relevant metrics such as accuracy, precision, recall, and the confusion matrix.
  - Result : The system now enables fast, reliable, and accurate classification of rural infrastructure projects into their correct PMGSY schemes, reducing manual workload and supporting better government decision-making.

SYSTEM APPROACH

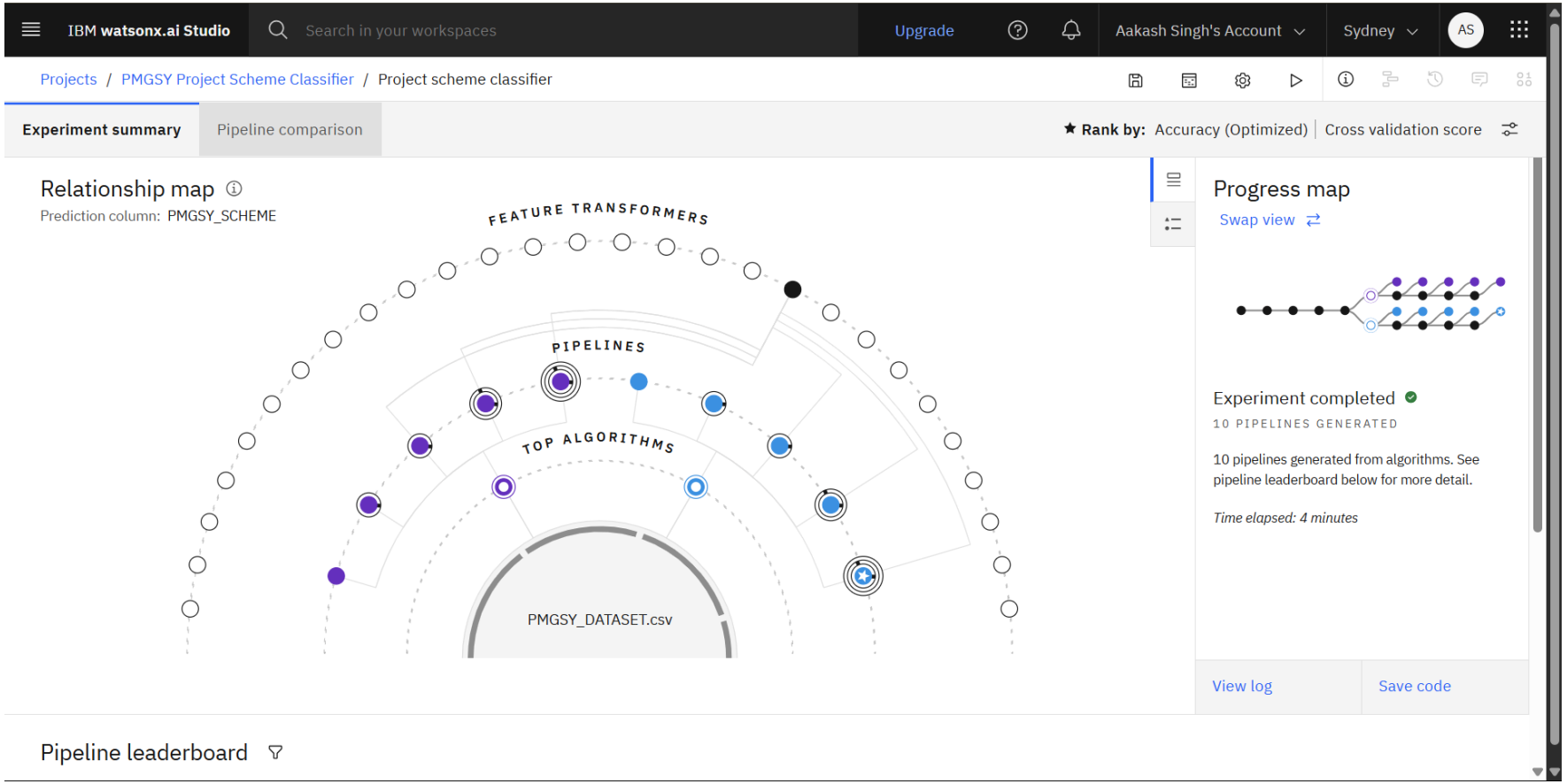
- **Platform:** IBM Cloud Lite services for scalable, cloud-based model development and deployment.
- **Tool:** IBM AutoAI for automated machine learning pipeline including data preprocessing, model training, and optimization.
- **Data Source:** AI Kosh dataset from the Government of India containing detailed road and bridge project data.
- **Programming Languages:** Python (for model exploration, evaluation, and deployment scripting).
- **Model Deployment:** Deployed trained model as an API endpoint on IBM Cloud for real-time project classification.
- **Additional Tools:** Jupyter notebooks for data analysis and visualization during development.

ALGORITHM & DEPLOYMENT

- **Algorithm:**
  Utilized IBM AutoAI for automated machine learning, which conducted feature selection, model training, optimization, and pipeline generation. The approach involved multiclass classification to predict the PMGSY scheme based on project features.
- **Model Training:**
  The AutoAI experiment was configured using the AI Kosh dataset, targeting the "PMGSY_SCHEME" column for classification. Training included holdout validation and scoring using accuracy.
- **Model Evaluation:**
  Pipelines were compared using accuracy and feature importance; the best pipeline was selected for deployment.
- **Deployment:**
  The final model was deployed on IBM Cloud Lite as a web service using IBM Watsonx.ai Runtime. It is accessible via API for real-time project classification.
- **Scoring & Prediction:**
  The deployed model accepts new project data, returning predicted PMGSY scheme along with probability scores, enabling accurate and quick scheme assignment.

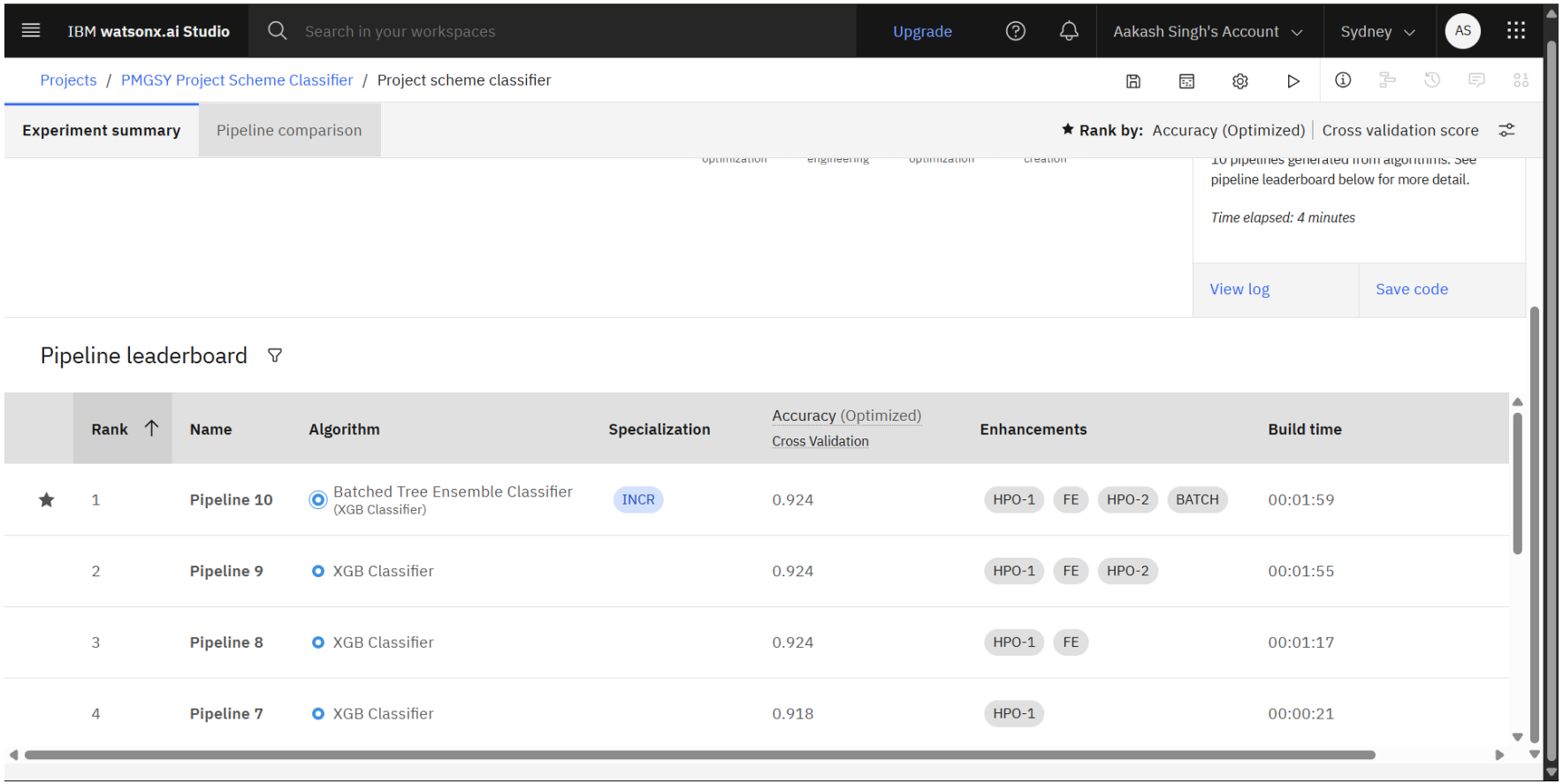
RESULT

- **Model Performance:** The top pipeline, using a Batched Tree Ensemble Classifier (XGB Classifier), achieved a cross-validated accuracy of 92.4%, making it highly reliable for classifying PMGSY project schemes.



RESULT

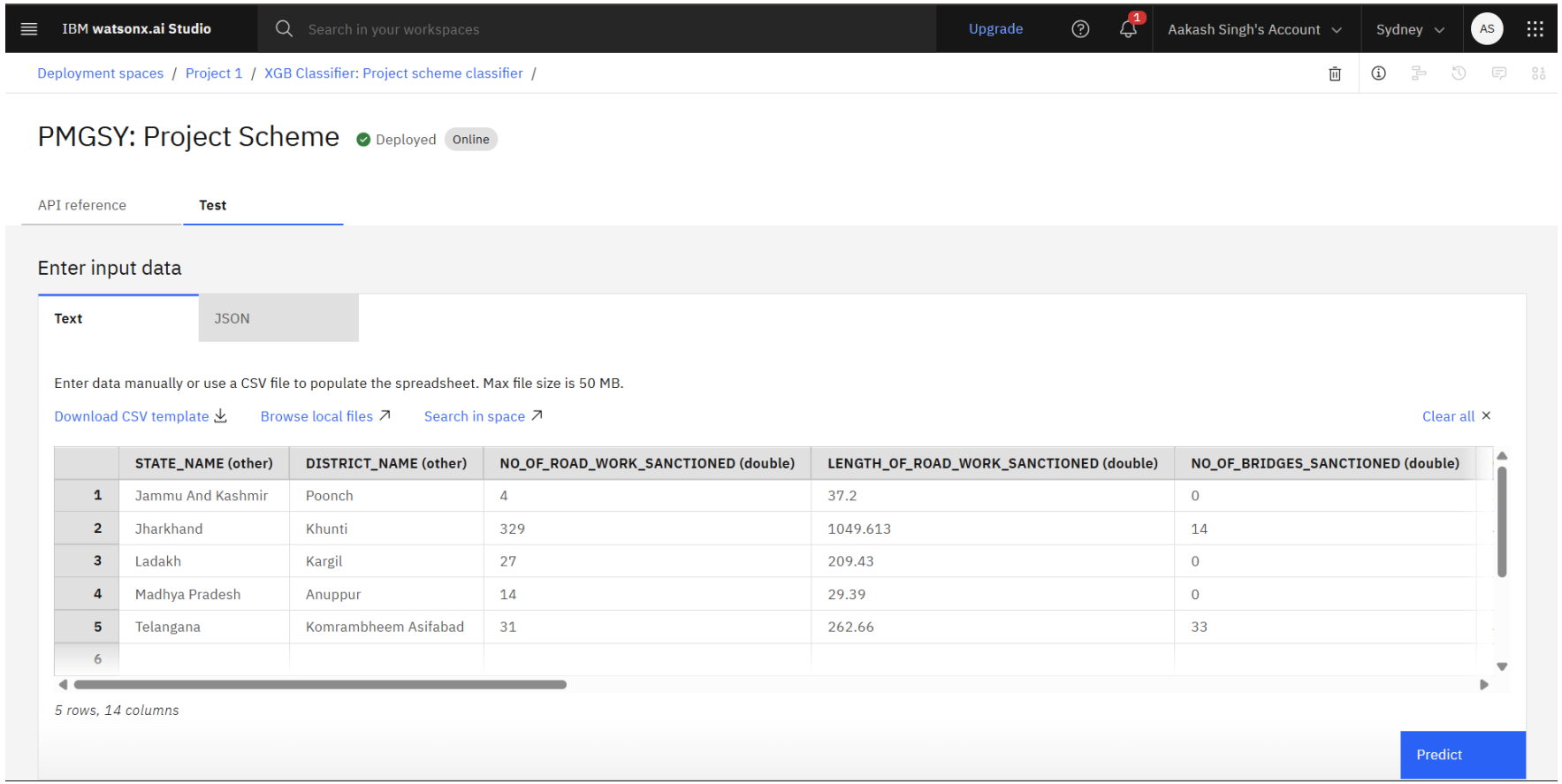
- **Pipeline Development:** AutoAI generated 10 different pipelines through data preprocessing, model selection, feature engineering, and ensemble methods. Pipeline 10 was selected as the best performer based on accuracy.



edunet foundation

RESULT

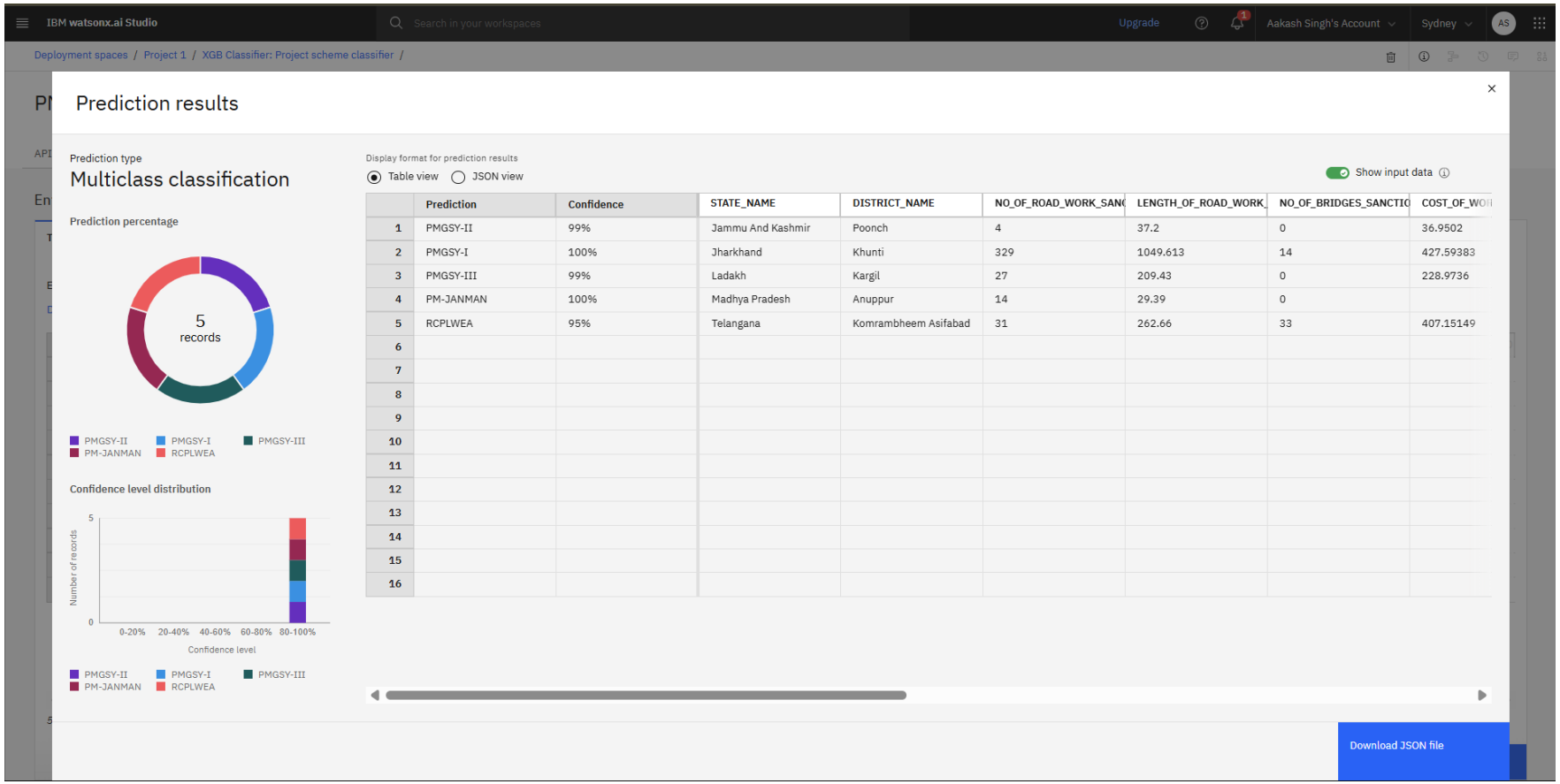
- **Deployment & Usage:** The final model was seamlessly deployed as an online service on IBM Cloud. Users can input new project data directly into the interface to receive instant scheme predictions.



edunet foundation

RESULT

- **Prediction Results:** On test data, the classifier confidently assigned schemes such as PMGSY-II, PMGSY-I, PMGSY-III, PM-JANMAN, and RCPLWEA with confidence levels between 95% and 100%. The system displayed both predicted classes and their associated probabilities for transparency.



	Prediction	Confidence	STATE_NAME	DISTRICT_NAME	NO_OF_ROAD_WORK_SANC	LENGTH_OF_ROAD_WORK_	NO_OF_BRIDGES_SANCTIO	COST_OF_WOI
1	PMGSY-II	99%	Jammu And Kashmir	Poonch	4	37.2	0	36.9502
2	PMGSY-I	100%	Jharkhand	Khunti	329	1049.613	14	427.59383
3	PMGSY-III	99%	Ladakh	Kargil	27	209.43	0	228.9736
4	PM-JANMAN	100%	Madhya Pradesh	Anuppur	14	29.39	0	
5	RCPLWEA	95%	Telangana	Komrambheem Asifabad	31	262.66	33	407.15149
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								

edunet foundation

CONCLUSION

- This project successfully automated the classification of rural road and bridge projects into relevant PMGSY schemes using IBM AutoAI and cloud deployment. The resulting model, with a high accuracy of 92.4%, enables government planners and stakeholders to make quick and reliable decisions while eliminating the errors, delays, and resource constraints of manual classification. Our solution demonstrates the impact of machine learning in public sector infrastructure management—boosting transparency, driving efficiency, and supporting scalable policy analysis. Through rigorous model evaluation and seamless integration, the system stands ready to streamline project monitoring and future scheme assessments.

FUTURE SCOPE

- **Expanded Data Integration:** Incorporate real-time project updates and additional data sources, such as satellite imagery and local weather, to further improve classification accuracy and adaptability.
- **Enhanced Algorithmic Approaches:** Explore advanced techniques (e.g., deep learning, ensemble meta-models) for even more robust project classification.
- **Geographical & Temporal Analysis:** Add modules for region-wise scheme trends, project progress over time, and predictive analytics for upcoming project needs.
- **Policy & Impact Assessment:** Use the system for deeper evaluation of scheme effectiveness, impact measurement, and data-driven policy recommendations.
- **User Experience:** Develop a more intuitive dashboard and APIs to allow government officials and planners to upload, visualize, and analyze project data easily.
- **Scalability:** Extend the solution framework to cover other government schemes or public infrastructure domains using similar AI-based classification methods.

REFERENCES

- AI Kosh Dataset: Pradhan Mantri Gram Sadak Yojana (PMGSY)
  - https://aikosh.indiaai.gov.in/home/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html
- IBM Watson Studio AutoAI Documentation
  - https://www.ibm.com/docs/en/watsonx/w-and-w/2.0.0?topic=models-autoai
- IBM watsonx.ai API Reference
  - https://cloud.ibm.com/apidocs/watsonx-ai
- PMGSY Official Guidelines and Publications
  - https://www.myscheme.gov.in/schemes/pmgsy
  - https://megcnrd.gov.in/forms/PMGSY.pdf
- Machine Learning Classification Best Practices
  - https://www.geeksforgeeks.org/machine-learning/getting-started-with-classification/
  - https://www.ibm.com/docs/en/watsonx/w-and-w/2.0.0?topic=models-build-model-autoai

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence
IBM SkillsBuild
IBM

Aakash Singh

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 15, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/96dd8b4c-f90e-4aa2-868f-41a79e8d2c66

IBM

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Aakash Singh

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

edunet foundation