IBM AICTE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECT IN ML

Presented By:
1.Name:- Soumya Ranjan Patra
2.College Name :- Institute of Technical Education and Research ,SOA
3.Department:- B.Tech - CSE

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

- PMGSY aims to provide all-weather road connectivity to unconnected rural habitations.
- The program has multiple schemes: PMGSY-I, PMGSY-II, RCPLWEA, etc., each with unique objectives and specifications.
- There are thousands of current and finished projects that have to be classified in order to be monitored and have a budget assigned to them.
- Manual classification is time-consuming, inaccurate and lacks the ability to scale.
- Demand of an automatic machine learning model which classifies projects in terms of physical and financial aspects.

PROPOSED SOLUTION

- The proposed system aims to use machine learning to automatically classify rural infrastructure projects into the correct PMGSY scheme, reducing manual effort, improving accuracy, enabling transparent budget allocation, and supporting data-driven policy decisions with scalable and explainable predictions. The solution will consist of the following components :
- **Data Collection:**
  - Gather project data from government records, PMGSY portals, and official reports.
  - Include both physical attributes (length, terrain, road type) and financial details (cost, expenditure, funding).
- **Data Preprocessing:**
  - Clean the dataset by handling missing values, removing duplicates, and correcting inconsistencies.
  - Encode categorical variables and scale numerical features for efficient model training.
- **Machine Learning Algorithm:**
  - Use classification algorithms like Random Forest and XGBoost to predict the correct PMGSY scheme.
  - Evaluate models using metrics such as accuracy, precision, recall, and F1-score to select the best performer.
- **Deployment:**
  - Deploy the trained model as an API or integrate it into a dashboard for real-time project classification.
  - Ensure accessibility for government officials with a user-friendly interface and secure data handling.
- **Evaluation:**
  - Assess model performance using accuracy, precision, recall, F1-score, and confusion matrix.
  - Perform cross-validation to ensure model reliability and generalization to new project data.

SYSTEM APPROACH

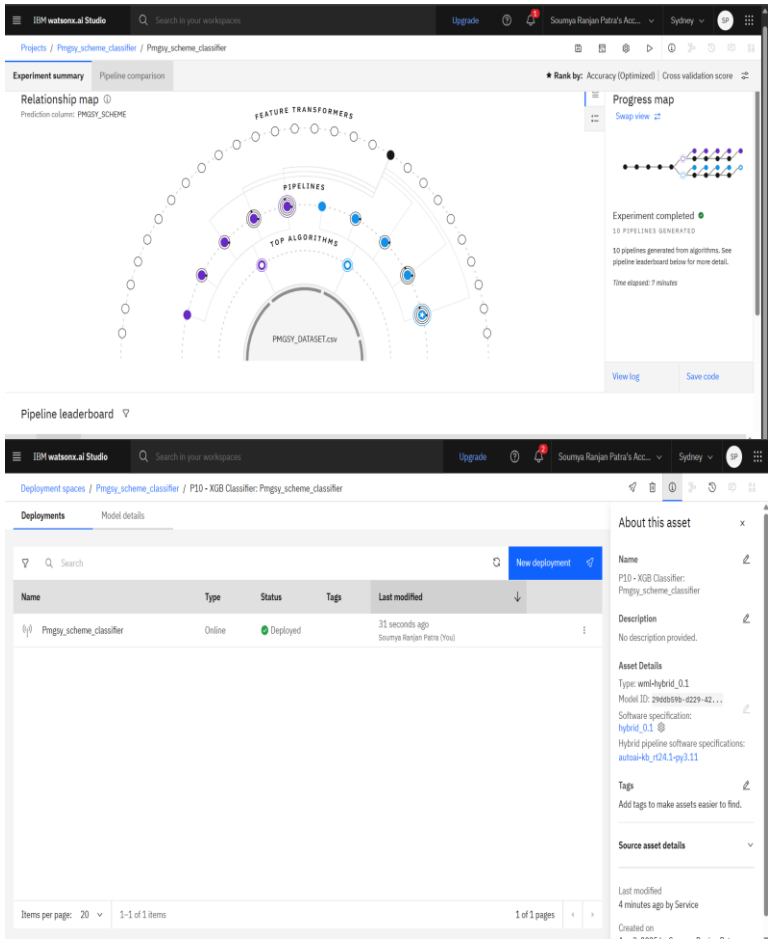
The System Approach section outlines the overall strategy and methodology for developing and implementing the PMGSY scheme classification system using machine learning.

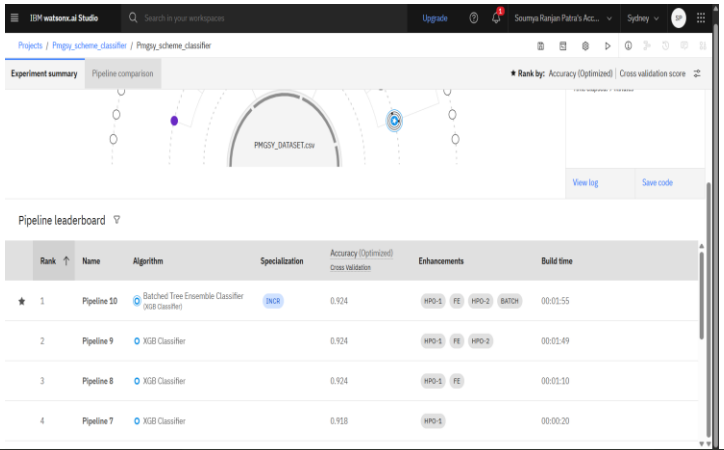
- **System requirements :**
  - Computer with minimum 8 GB RAM and multi-core processor
  - Python (version 3.8 or above)
  - Dataset containing physical and financial project attributes
  - Internet access for library installation and dataset updates
- **Library required to build the model :**
  - Pandas – for data loading and preprocessing
  - Numpy – for numerical computations
  - scikit-learn – for model building, training, and evaluation
  - xgboost / lightgbm – for gradient boosting classification
  - matplotlib / seaborn – for visualization of results
  - shap – for explainable AI insights

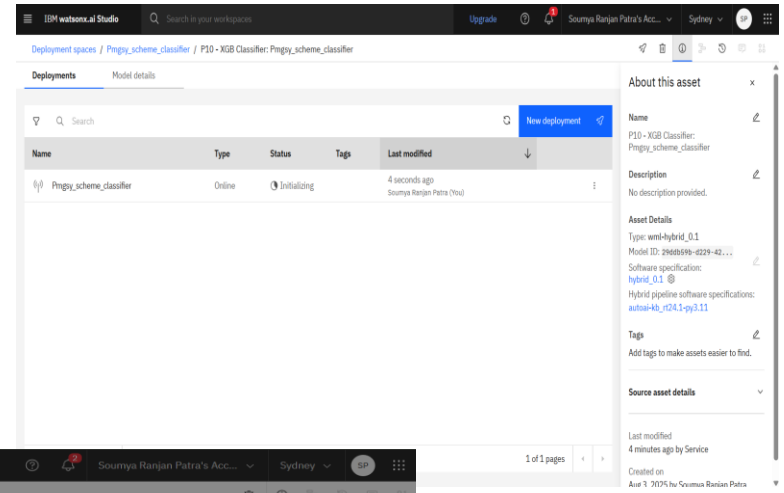
ALGORITHM & DEPLOYMENT

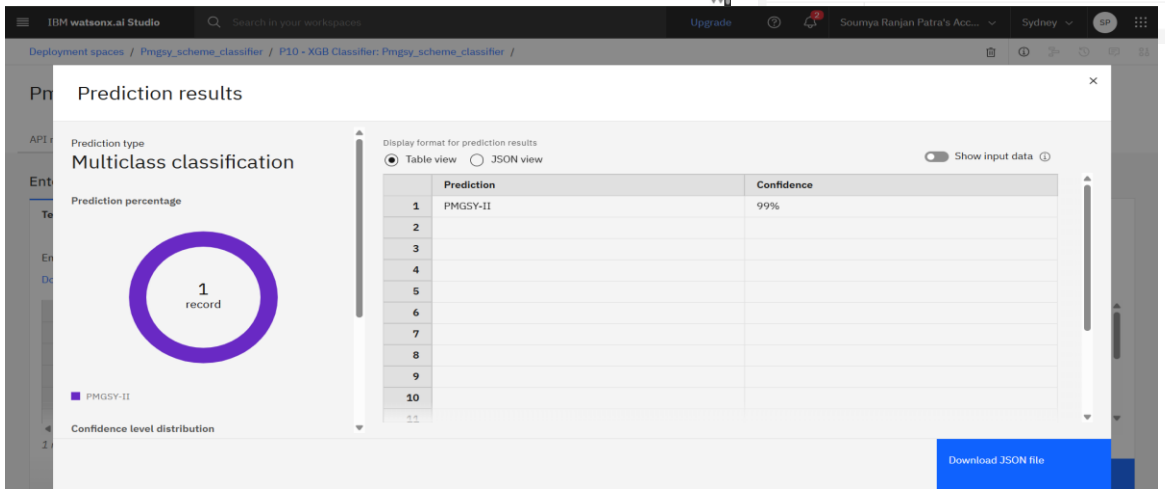
- **Algorithm Selection:**
  - Chosen Algorithm: XGBoost Classifier
  - Handles both numerical and categorical features with high efficiency.
  - Built-in regularization helps reduce overfitting and improve generalization.
  - Provides feature importance scores for explainable and transparent predictions.
- **Data Input:**
  - Physical Attributes: Road/bridge length, terrain type, road surface type, traffic estimates, and project location.
  - Financial Attributes: Sanctioned project cost, actual expenditure, funding source breakdown, and year of sanction.
  - Project Specifications: Type of construction (road/bridge), connectivity priority, and scheme-specific guidelines.
- **Training Process:**
  - Trained the XGBoost Classifier on historical PMGSY project data with physical and financial attributes, using stratified train-test split.
  - Applied cross-validation and hyperparameter tuning to optimize model performance and prevent overfitting.
  - Evaluated the model using accuracy, precision, recall, and F1-score to select the best configuration.
- **Prediction Process:**
  - The trained XGBoost Classifier takes new project data with physical and financial attributes as input.
  - Processes the input features through the model to predict the most probable PMGSY scheme.
  - Can be integrated with a real-time dashboard or API for instant classification of incoming project data.

RESULT









edunet foundation

CONCLUSION

- Developed an XGBoost-based system to automatically classify PMGSY projects using physical and financial attributes.
- Achieved high accuracy and reliability, improving transparency and efficiency in project categorization. Addressed challenges like missing values, imbalanced data, and hyperparameter tuning through preprocessing and optimization techniques.
- Demonstrated the importance of accurate classification for proper budget allocation, monitoring, and rural infrastructure planning.
- System has potential for scaling nationwide with further enhancements and integration of additional data sources.

FUTURE SCOPE

- Incorporate additional data sources such as GIS data, satellite imagery, and real-time project progress updates to improve prediction accuracy.
- Optimize the XGBoost algorithm using advanced hyperparameter tuning methods like Bayesian optimization for better performance.
- Expand system coverage to classify infrastructure projects across multiple states or regions, adapting to regional variations in schemes and project attributes.
- Integrate emerging technologies like edge computing for on-site, low-latency classification and decision-making in rural areas with limited connectivity.
- Explore advanced machine learning techniques such as deep neural networks and ensemble models for further improving classification accuracy.
- Implement continuous learning pipelines to update the model automatically as new project data becomes available.

REFERENCES

1. T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
2. F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
3. Ministry of Rural Development, Government of India, “Pradhan Mantri Gram Sadak Yojana – Programme Guidelines and Project Data,” [Online]. Available: https://pmgsy.nic.in/
4. J. Brownlee, *Data Preparation for Machine Learning: Data Cleaning, Feature Selection, and Data Transforms in Python*. Machine Learning Mastery, 2020.
5. S. M. Lundberg and S. I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Advances in Neural Information Processing Systems (NIPS)*, 2017.

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Soumya Ranjan Patra

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 19, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/cfeb1078-8c62-42f0-9077-408f56a0d718

IBM

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Journey to Cloud: Envisioning Your Solution

IBM SkillsBuild

IBM

Soumya Ranjan Patra

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/74c4bf2e-5958-45b1-825c-a7d429e96d32

IBM

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Soumya Ranjan Patra

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 31 Jul 2025 (GMT) **Learning hours:** 20 mins

edunet foundation

GITHUB REPOSITORY LINK

GitHub Link :- https://github.com/Soumyapatra-0505

THANK YOU

edunet foundation