PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:
Jayasree S
B.Tech AI&DS(4th year)
Anna University Regional Campus,Coimbatore

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

Intelligent Classification of Rural Infrastructure Projects

Design and evaluate a machine learning model to automatically classify PMGSY road and bridge projects into their respective schemes (PMGSY-I, PMGSY-II, RCPLWEA, etc.) based on physical and financial attributes, replacing manual, error-prone categorization for scalable analysis and decision-making.

PROPOSED SOLUTION

- The goal is to develop a machine learning-based classification system to automatically categorize PMGSY road and bridge construction projects into their respective schemes (PMGSY-I, PMGSY-II, RCPLWEA, etc.) using physical and financial features. This solution enhances transparency, speeds up analysis, and reduces manual errors in large-scale project monitoring.
- **Data Collection:**
- **Gather structured tabular data from the AI Kosh PMGSY dataset, including number and length of roads/bridges sanctioned and completed.**
- **Source metadata such as state, district, and region for contextual relevance.**

Data Preprocessing:

- **Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.**
- **Apply feature engineering to extract meaningful predictors (e.g., % completion, cost per km, etc.).**

- **Machine Learning Algorithm:**
  - **Use XGBoost Classifier (XGBClassifier) for its high performance in structured/tabular data.**
  - **Perform:Hyperparameter optimization (HPO-1, HPO-2),Feature engineering (FE),Compare multiple pipelines for best accuracy.**
- **Deployment:**
  - **Integrate the trained model into a cloud-based interface using IBM Watsonx.ai and Ensure support for batch-based classification.**
  - **Evaluation:**
  - **Evaluate performance using Accuracy, Precision, Recall, F1-score,Cross-validation for reliability**
  - **Best-performing model:Pipeline 8 (XGBClassifier) with 92.4% accuracy after enhancements**
  - **Result:**
  - **Reduces manual errors and delays in classification. Enhances transparency and speeds up analysis for large-scale rural infrastructure monitoring**

edunet foundation

SYSTEM APPROACH

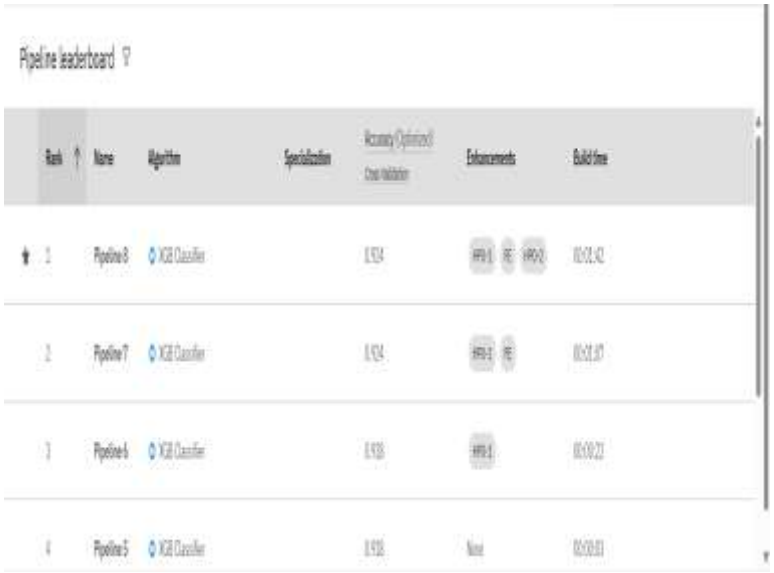
System requirements	Libraries and Tools Used
Operating System: Windows 10/11 (via WSL) or Linux	pandas: For data preprocessing and analysis
Environment: IBM Watson Studio (Cloud-based)	numpy: For numerical computation
Python Version: 3.8 or later	matplotlib & seaborn: For data visualization
CPU/GPU: Intel i5/i7 or equivalent, GPU optional for faster model training	scikit-learn: For traditional ML models
RAM: 8GB minimum (16GB recommended)	xgboost: For gradient boosting classifier ibm-watson-machine-learning: For AutoAI pipeline creation and deployment

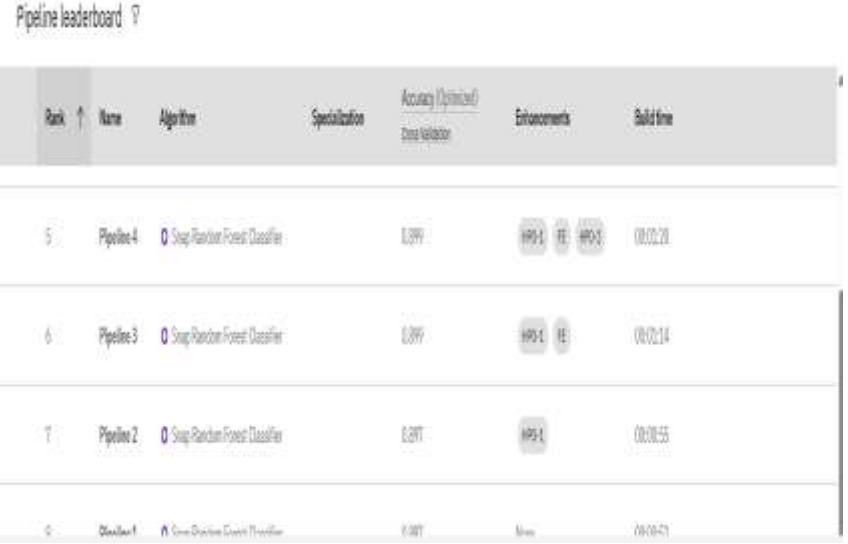
edunet foundation

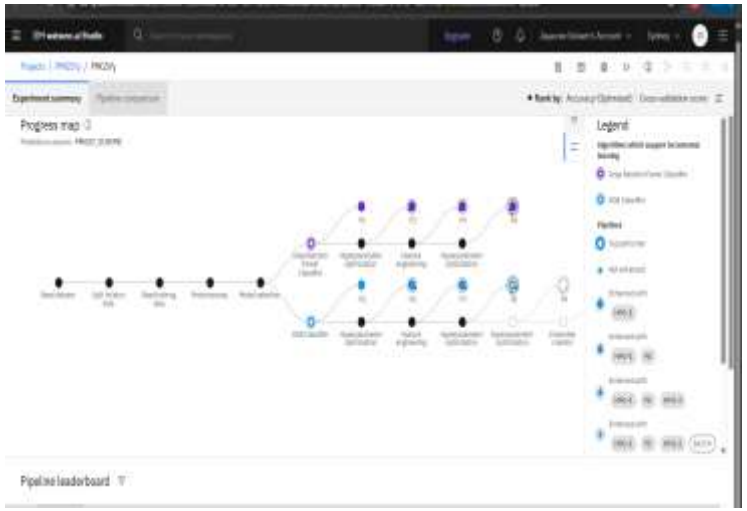
ALGORITHM & DEPLOYMENT

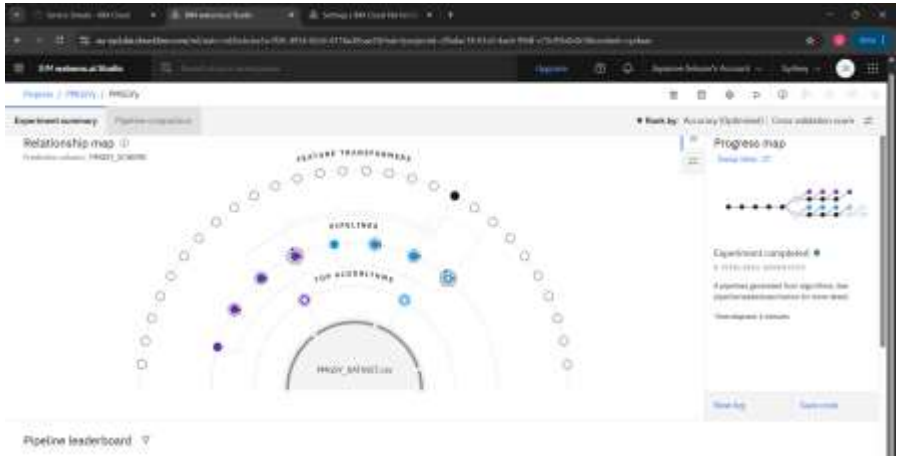
- **Algorithm Selection:**
  - XGBoost, KNN, and Random Forest classifiers were evaluated using IBM Watson AutoAI. XGBoost was selected for its robustness and accuracy on structured PMGSY data..
- **Data Input:**
  - Features included project cost, % completion, road/bridge lengths, and geographic metadata like state and district..
- **Training Process:**
  - AutoAI performed preprocessing, feature engineering, and hyperparameter tuning. Models were validated using cross-validation.
  - **Prediction Process:**
  - The best pipeline predicted the scheme (PMGSY-I, II, RCPLWEA, etc.) for new projects with over 91% accuracy..

RESULT









edunet foundation

CONCLUSION

- The proposed machine learning solution effectively classified PMGSY infrastructure projects with over 91% accuracy, reducing manual errors and saving time. AutoAI pipelines with enhancements like feature engineering and hyperparameter tuning improved model performance. Challenges like missing data were addressed through preprocessing. The system improves scalability and decision-making in rural project analysis.

FUTURE SCOPE

- The system can be enhanced by incorporating additional datasets such as satellite imagery, GIS data, or socio-economic indicators to improve classification accuracy. Future work can explore advanced algorithms like deep learning or ensemble stacking for better performance. Expanding the system to support multiple states or nationwide deployment is feasible. Integration with edge computing can enable real-time insights at the field level.

REFERENCES

- Ministry of Rural Development, Government of India – **PMGSY Scheme Guidelines** https://pmgsy.nic.in
- IBM Watson Studio Documentation – **AutoAI and Machine Learning Services** https://www.ibm.com/cloud/watson-studio
- 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.*
- Géron, A. (2019). **Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow**. 2nd Edition, O'Reilly Media.
- Scikit-learn Documentation – **Machine Learning in Python** https://scikit-learn.org
- AI Kosh Open Data Platform – **Rural Infrastructure Project Dataset** https://data.gov.in

IBM CERTIFICATIONS

Getting Started with Artificial Intelligence

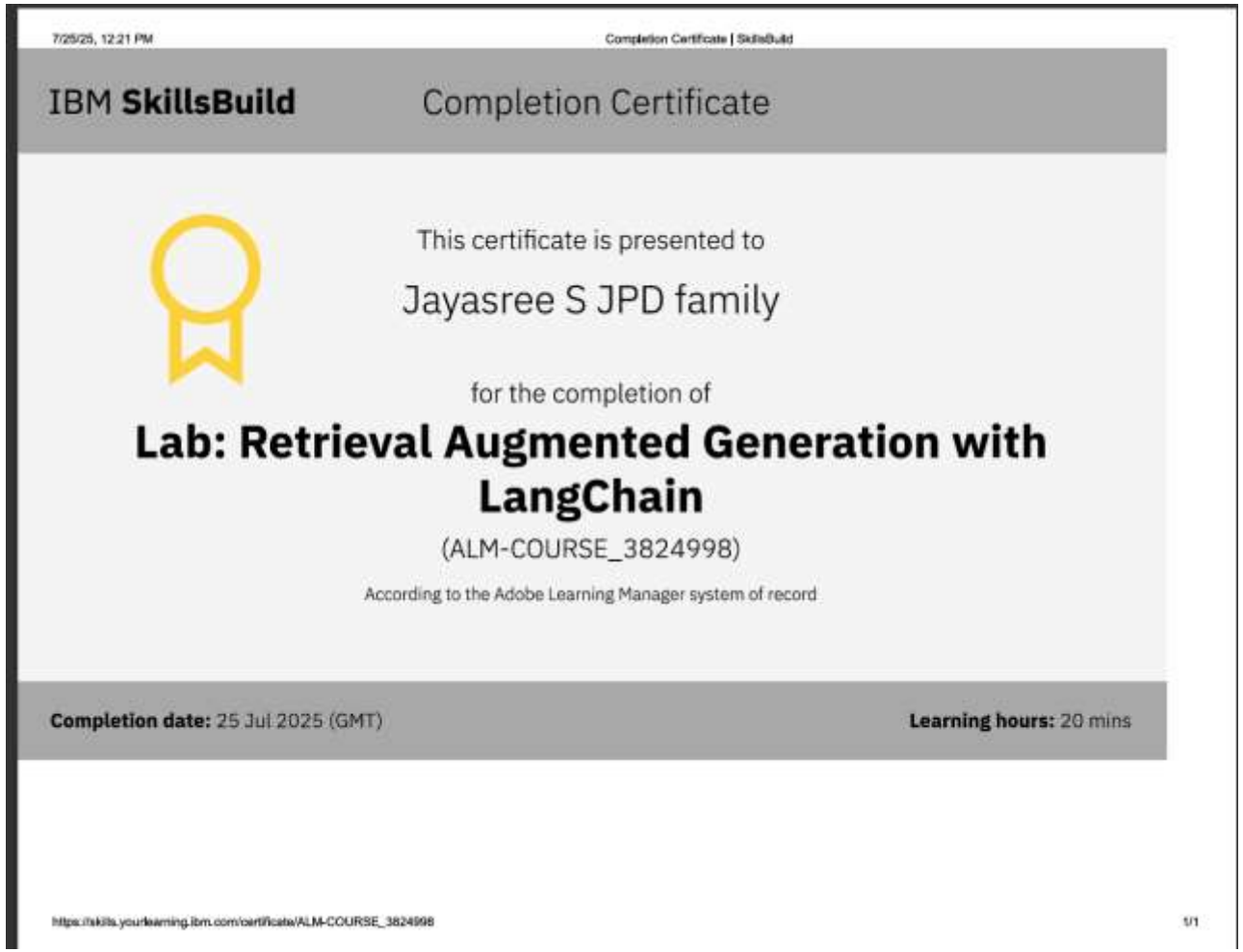
IBM **SkillsBuild**

IBM

edunet foundation

IBM CERTIFICATIONS

IBM CERTIFICATIONS



7/25/25, 12:21 PM Completion Certificate | SkillsBuild

IBM **SkillsBuild**

Completion Certificate

This certificate is presented to

Jayasree S JPD family

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 25 Jul 2025 (GMT)

Learning hours: 20 mins

https://skills.yourlearning.ibm.com/certificate/ALM-COURSE_3824998 1/1

THANK YOU

edunet foundation