CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:
1. Aditya Raj – Kalinga University Raipur –
B.Tech in Computer Science Engineering

edunet foundation

OUTLINE

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

- **The proposed system aims to address the challenge of automatically classifying rural infrastructure projects under the correct PMGSY scheme. This is achieved by leveraging data analytics and machine learning techniques to automate and improve the accuracy of classification, enabling better monitoring, funding transparency, and impact analysis. The solution consists of the following components:**
- **Data Collection:**
  - **Source historical data from the AI Kosh PMGSY dataset, which includes various attributes related to road and bridge construction projects.**
  - **Collect both physical and financial parameters such as project cost, length, type (road or bridge), and implementation dates for accurate labeling under PMGSY schemes (PMGSY-I, PMGSY-II, RCPLWEA, etc.).**
- **Data Preprocessing:**
  - **Clean and preprocess the dataset to manage missing values, remove duplicates, and normalize feature scales.**
  - **Perform feature engineering to derive meaningful inputs such as cost per kilometer, project duration, and funding allocation patterns that influence scheme classification.**
  - **Encode categorical variables and split data into training and testing sets.**
- **Machine Learning Algorithm:**
  - **Utilize *AutoAI* within *IBM Watsonx.ai Studio* to automate the training and optimization of machine learning pipelines.**
  - **The chosen algorithm is a *Batched Tree Ensemble Classifier (XGB Classifier),* known for high performance on tabular data.**
  - **The model achieved an *accuracy of 0.924*, demonstrating strong predictive capability.**
  - **The pipeline (named *autoai-kb_rt24.1-py3.11*) supports incremental training, making it suitable for large and evolving datasets.**

PROPOSED SOLUTION

- **Deployment:**
  - **The trained model was deployed using IBM Cloud Lite services under the model *asset ID: 06ac3118-8249-4e45-803c-4fcd42a10d2c*.**
  - **The deployment allows stakeholders to input project details and receive instant classification suggestions under the correct PMGSY scheme.**
  - **Future plans include global hosting for broader access and scalability.**
- **Evaluation:**
  - **Model evaluation was conducted using a dedicated test dataset.**
  - **The final classification accuracy was 92.4%, indicating high reliability.**
  - **Continuous monitoring and feedback will be used to retrain and fine-tune the model as new data becomes available.**

Result:

The solution successfully automates the classification of rural infrastructure projects with high accuracy, reducing manual errors and improving operational efficiency for planners and policy analysts. It demonstrates the potential of AI-driven solutions in supporting large-scale government programs through intelligent data use and machine learning.

SYSTEM APPROACH

System Requirements:

- **Platform:** IBM Cloud Lite
- **Development Environment:** Watsonx.ai Studio (AutoAI)
- **OS:** Cloud-based, accessible via any modern browser
- **Hardware:** No specific hardware needed—cloud-based resources used

Libraries Required:

The hybrid *AutoAI pipeline* (***autoai-kb_rt24.1-py3.11***) utilizes the following key libraries:

- **autoai-libs (v2.0.0):** For AutoAI automation and pipeline management
- **lale (v0.8.0):** Integration of AutoAI pipelines with scikit-learn
- **lightgbm (v4.2.0):** Fast gradient boosting framework
- **xgboost (v2.0.3):** Used for the final Batched Tree Ensemble Classifier
- **numpy (v1.26.4):** For numerical computations
- **pandas (v2.1.4):** Data preprocessing and handling
- **scikit-learn (v1.3.0):** ML tools for evaluation and preprocessing
- **scipy (v1.11.4):** Scientific computations
- **snapml (v1.14.5):** IBM's accelerated ML library for large-scale training

ALGORITHM & DEPLOYMENT

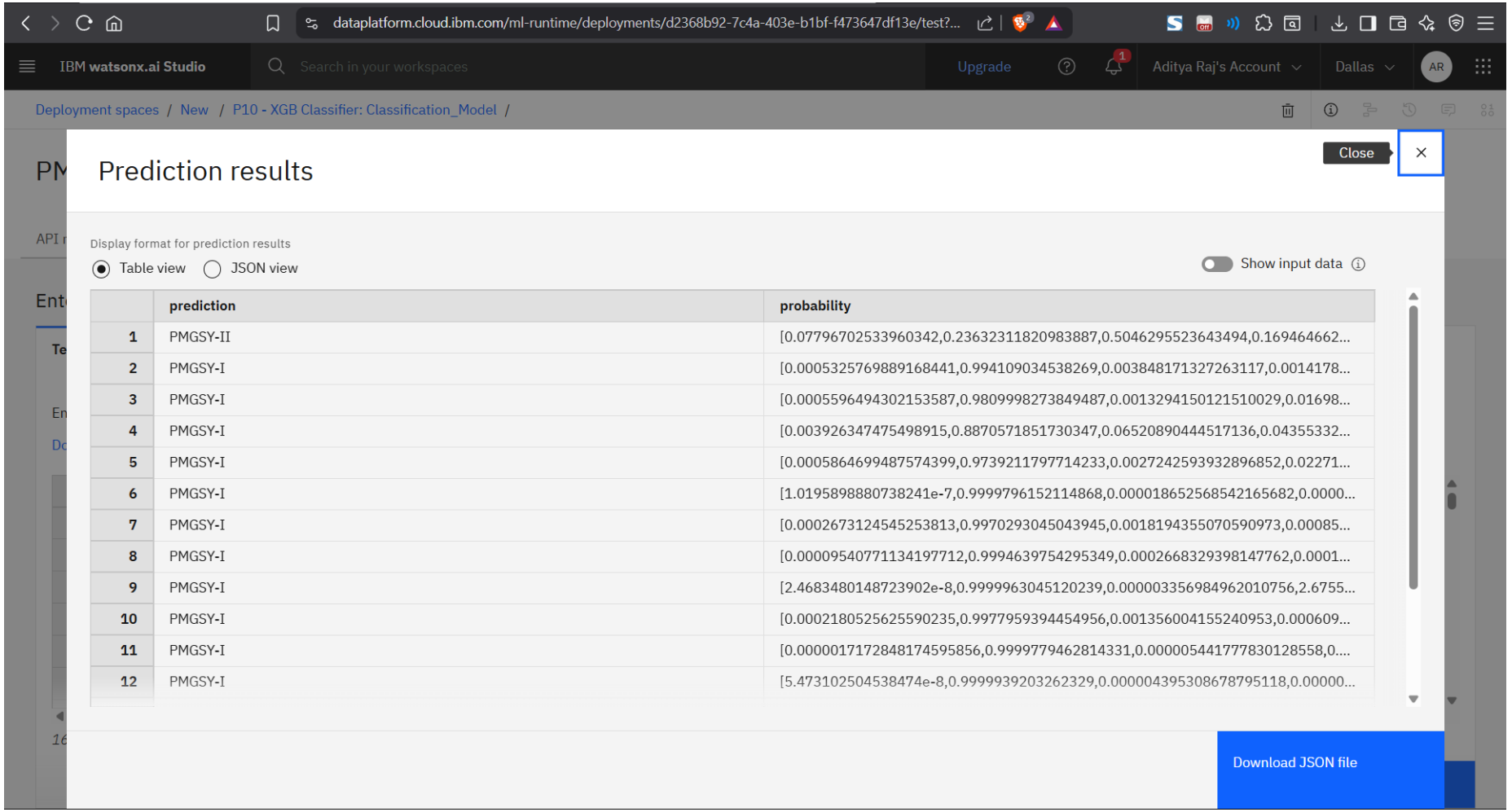
Algorithm:

- **Algorithm Selection:**
  Used **XGBoost Classifier** via AutoAI for its high accuracy and ability to handle structured tabular data.
- **Data Input:**
  Model takes project attributes like cost, length, type, and funding details as input features.
- **Training Process:**
  AutoAI handled preprocessing, feature selection, and hyperparameter tuning with cross-validation.
- **Prediction Process:**
  Model predicts the appropriate PMGSY scheme based on input features with **92.4% accuracy**.

Deployment:

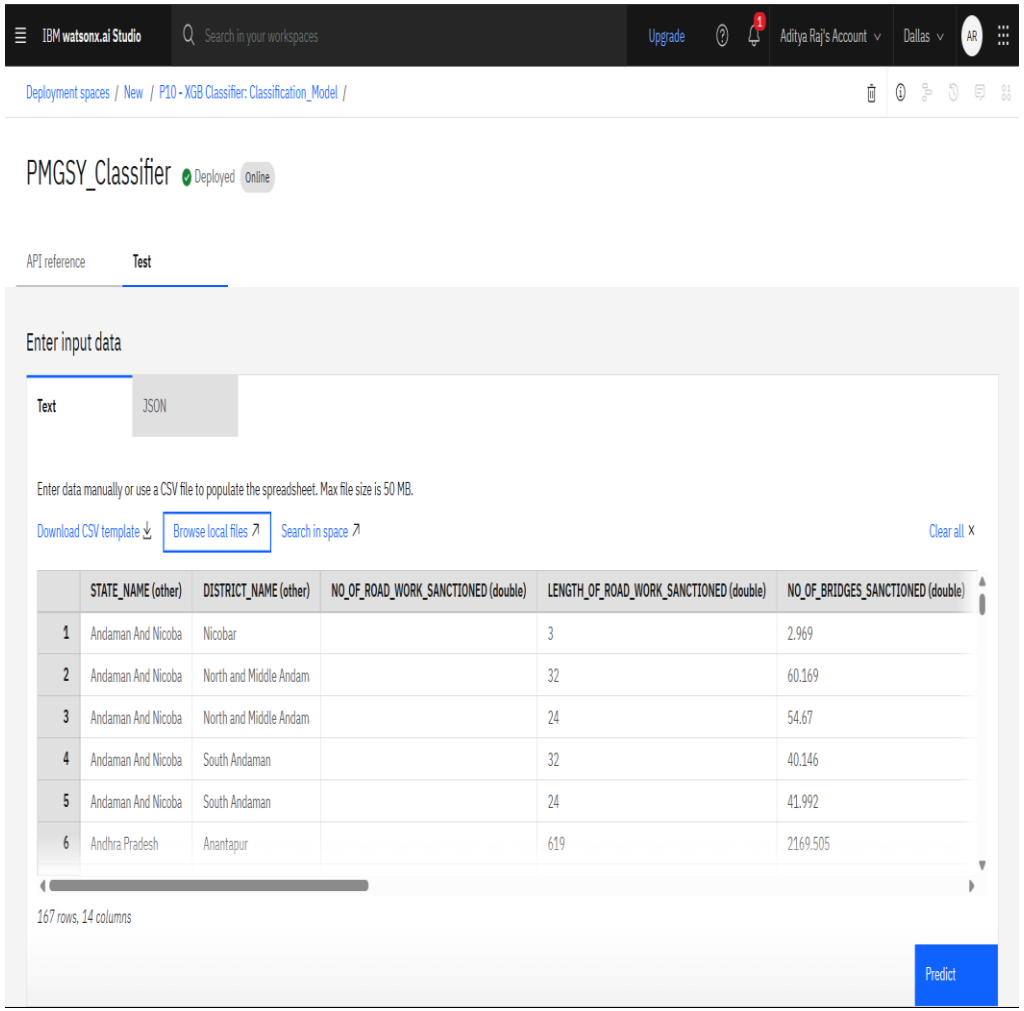
- **Deployment Strategy:**
  Model deployed on **IBM Cloud Lite** using Watson Machine Learning services.
  The model asset is identified by ID : *06ac3118-8249-4e45-803c-4fcd42a10d2c*
- **Access and Integration:**
  Users can input project data and receive real-time PMGSY scheme predictions via API or UI.
- **Scalability and Future Hosting:**
  Solution supports scale and will be **globally hosted** for wider use in public infrastructure systems.

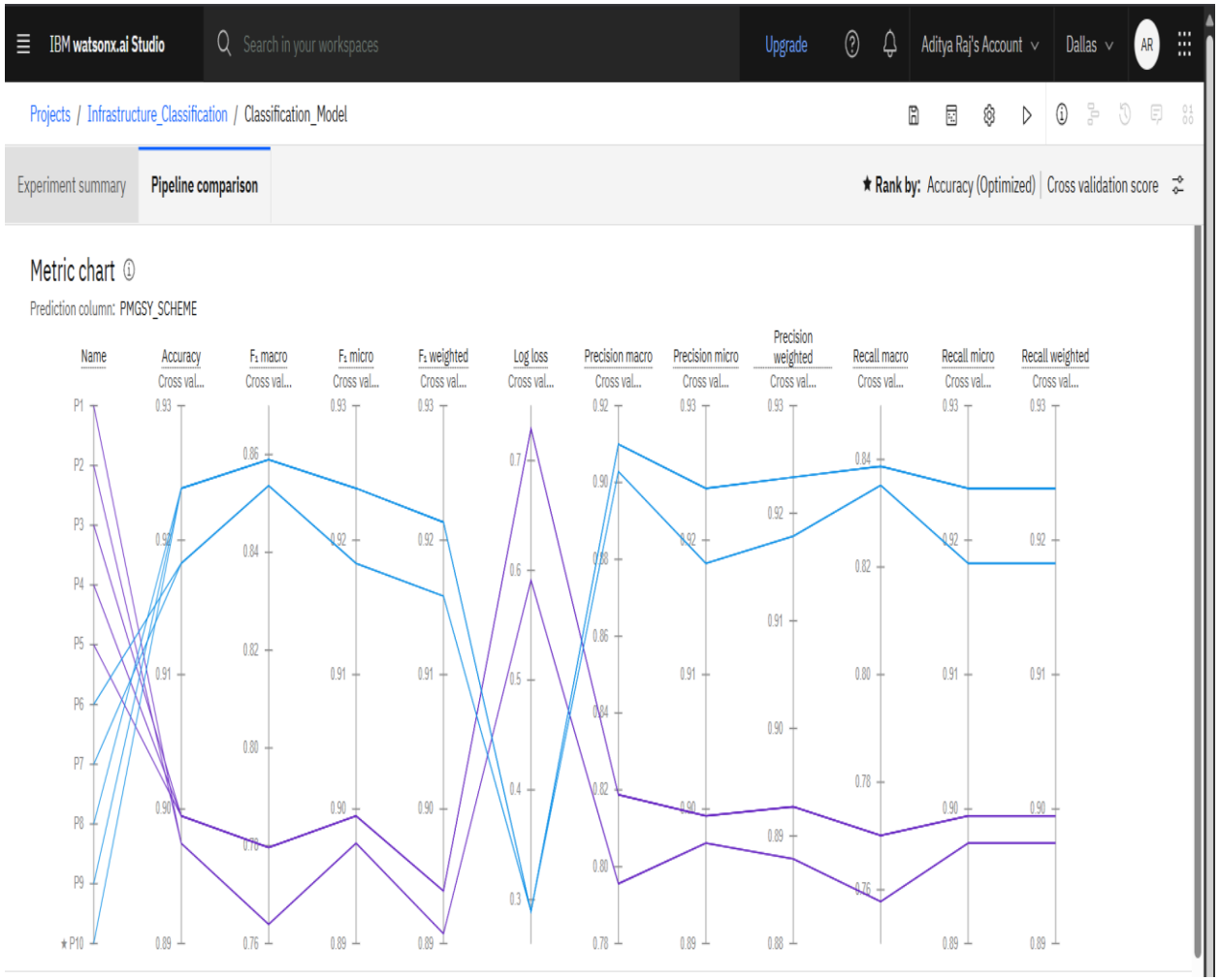
RESULT



	prediction	probability
1	PMGSY-II	[0.07796702533960342,0.23632311820983887,0.5046295523643494,0.169464662...
2	PMGSY-I	[0.0005325769889168441,0.994109034538269,0.003848171327263117,0.0014178...
3	PMGSY-I	[0.0005596494302153587,0.9809998273849487,0.0013294150121510029,0.01698...
4	PMGSY-I	[0.003926347475498915,0.8870571851730347,0.06520890444517136,0.04355332...
5	PMGSY-I	[0.0005864699487574399,0.9739211797714233,0.002724259393289685 2,0.02271...
6	PMGSY-I	[1.0195898880738241e-7,0.9999796152114868,0.0000186525685421656 82,0.0000...
7	PMGSY-I	[0.0002673124545253813,0.9970293045043945,0.0018194355070590973,0.00085...
8	PMGSY-I	[0.0000954077113419771 2,0.9994639754295349,0.0002668329398147762,0.0001...
9	PMGSY-I	[2.4683480148723902e-8,0.9999963045120239,0.0000033569849620107 56,2.6755...
10	PMGSY-I	[0.0002180525625590235,0.9977959394454956,0.0013560041552409 53,0.000609...
11	PMGSY-I	[0.0000017172848174595856,0.9999779462814331,0.0000054417778301285 58,0....
12	PMGSY-I	[5.473102504538474e-8,0.9999939203262329,0.0000043953086787951 18,0.00000...

RESULT





CONCLUSION

- Built an ML model to classify rural projects under correct PMGSY schemes.
- Achieved high accuracy of 92.4% using XGBoost via Watsonx.ai AutoAI.
- Automated classification improves efficiency and reduces manual errors.
- **Key challenge**: handling imbalanced data and complex financial features.
- Resolved using AutoAI's automated preprocessing and model tuning.
- **Future improvement**: add real-time updates and geospatial features.
- Supports better planning, budget allocation, and impact analysis.
- Demonstrates AI's role in enhancing large-scale government programs.

Github Repository Link :-

github.com/adityarajX/Edunet_IBM_internship

FUTURE SCOPE

- Integrate additional data sources like satellite imagery and real-time project updates.
- Expand the system to cover multiple states and rural development schemes.
- Optimize the model using advanced algorithms like ensemble deep learning or transformers.
- Incorporate geospatial analytics for location-aware classification.
- Enable incremental learning to adapt continuously with new project data.
- Explore edge computing for faster, on-site predictions in remote areas.
- Develop a dashboard or mobile interface for easier access by government officials.
- Ensure multi-language support for broader accessibility in rural regions.

REFERENCES

- **AI Kosh PMGSY Dataset**

Source of project data used for model training and evaluation.
https://aikosh.indiaai.gov.in

- **IBM Watsonx.ai Documentation**

Used for building, training, and deploying the AutoAI model on IBM Cloud.
https://www.ibm.com/products/watsonx-ai

- **IBM Watsonx.ai Studio (Lite Plan)**

Platform used to train, optimize, and deploy the ML model in a cloud environment.
https://www.ibm.com/products/watsonx-ai

- **IBM Watson Machine Learning (WML) Runtime**

Service used for deploying and serving the trained classification model.
https://cloud.ibm.com/catalog/services/watsonxai-runtime

IBM CERTIFICATIONS

IBM CERTIFICATIONS

IBM CERTIFICATIONS

7/24/25, 2:48 PM Completion Certificate | SkillsBuild

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Aditya Raj

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

edunet foundation

THANK YOU

edunet foundation