CAPSTONE PROJECT

INTELLIGENT CLASSIFICATION OF RURAL INFRASTRUCTURE PROJECTS

Presented By:

1. Shinjoy Sen-MCKV Institute of Engineering-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

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.

PROPOSED SOLUTION

The proposed solution aims 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. The solution will consist of the following components**:**

- **Data Collection:**
  - Gather historical data on Infrastructure Projects, including no of road-work, no of bridges, expenditure and other relevant factors for a certain PMGSY_SCHEME .
- **Data Preprocessing:**
  - Clean and preprocess the collected data to handle missing values, outliers, and inconsistencies.
  - Feature engineering to extract relevant features from the data that might impact PMGSY_SCHEME prediction.
- **Machine Learning Algorithm:**
  - Implement a machine learning algorithm, such as Random forest classifier, XGB classifier to predict PMGSY_SCHEME based on historical patterns.
  - Tune the model by Hyperparameter optimization, Feature engineering to enhance accuracy.
- **Deployment:**
  - Deploy our model which have higher accuracy on the IBM Cloud platform where we can predict PMGSY_SCHEME by providing the physical and financial characteristics.
- **Evaluation:**
  - Assess the model's performance using appropriate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or other relevant metrics.
  - Fine-tune the model based on feedback and continuous monitoring of prediction accuracy.

SYSTEM APPROACH

System requirements

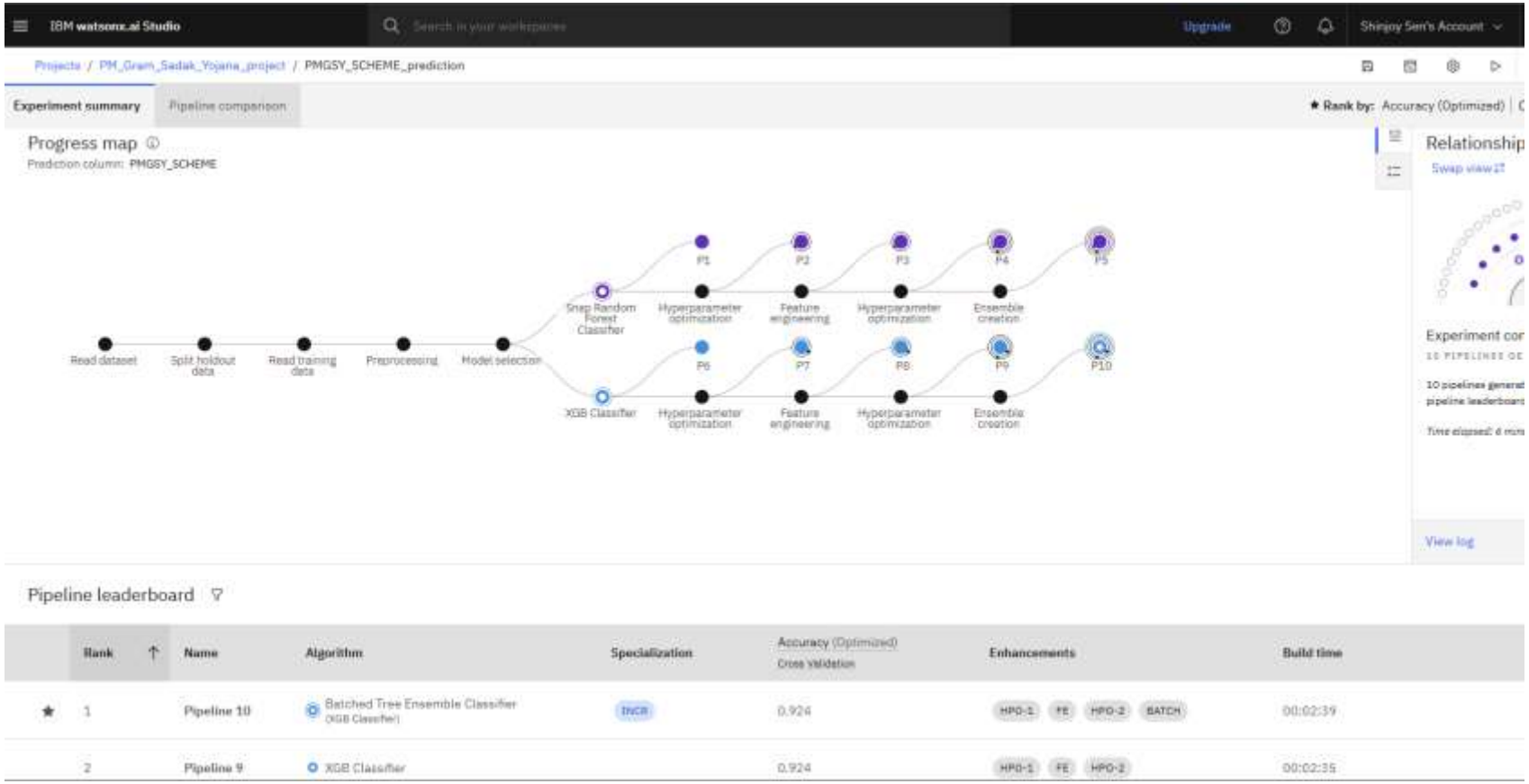
1. A Machine with min 8 GB RAM, 1vCPU

- **Library required to build the model**
  - **Data Handling & Preprocessing** : pandas, numpy, scikit-learn etc.
  - **Modeling:** xgboost, sklearn.ensemble.RandomForestClassifier etc.
  - **Fine-tuning:**
    - hyperopt for hyperparameter optimization.
    - sklearn.ensemble etc.

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  - Based on the prediction result algorithm was chosen random-forest classifier, XGB boost classifier in different pipelines. Among these XGB classifier has better accuracy after specific fine tuning like Hyperparameter optimization, ensemble methods etc.
- **Data Input:**
  - This model takes input parameters as no of road-work, no of bridges, expenditure and other relevant factors to predict PMGSY_SCHEME .
- **Training Process:**
  - The model is trained using provided data as 90% training data and 10% holding for testing.
  - After training for increase the accuracy we use Hyperparameter optimization, Feature Engineering and other fine tuning methods.
- **Prediction Process:**
  - After model trained and build we deploy our model in the cloud platform to test or predict result(PMGSY_SCHEME) by provided input features like no of road-work, no of bridges, expenditure etc physical and financial characteristics.

RESULT

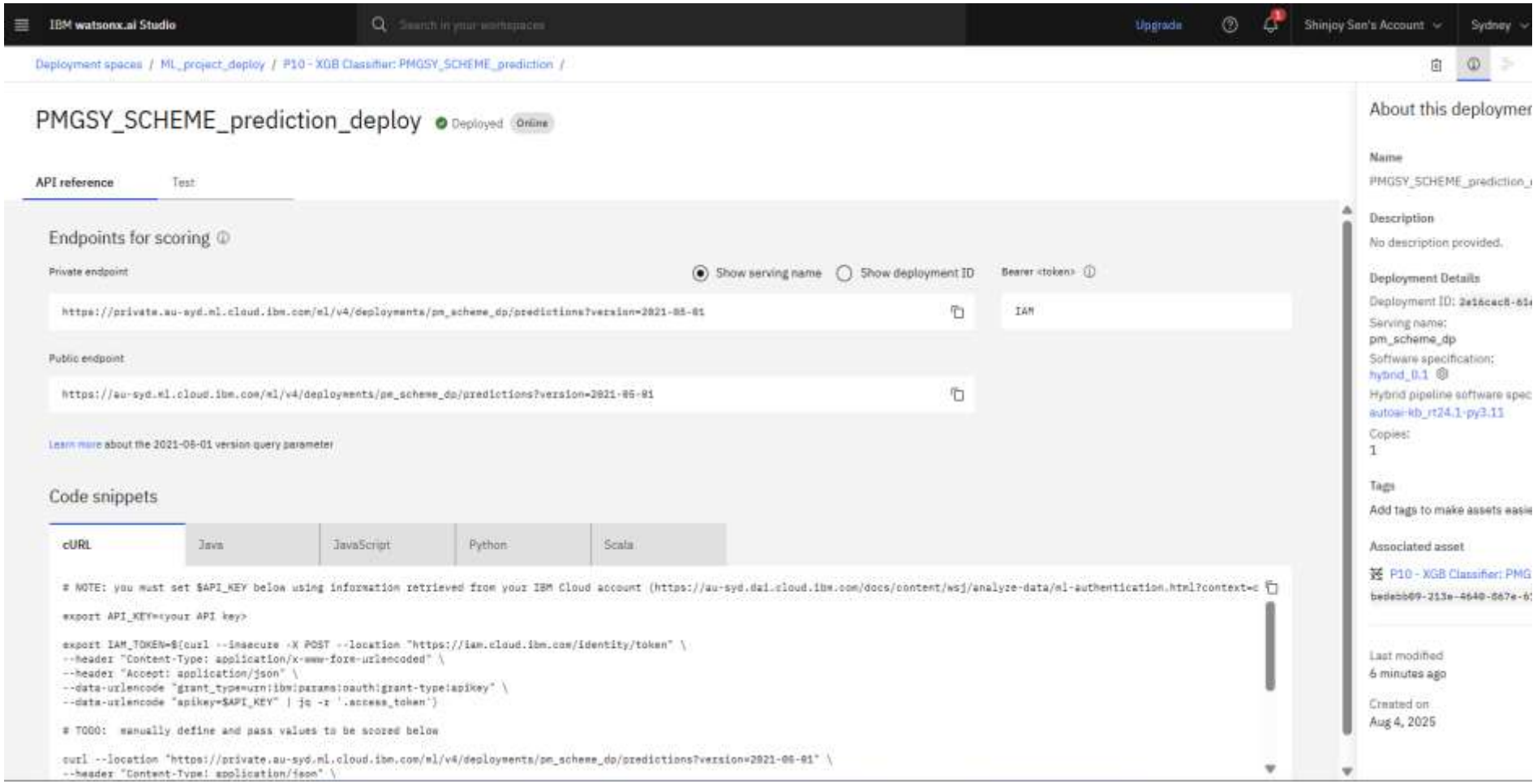


Here we build our model in IBM cloud platform using watsonx.ai service.

In multiple pipelines the model is tuning and checking for accuracy. Then we select the pipeline which have max accuracy. Here pipeline 10.

edunet foundation

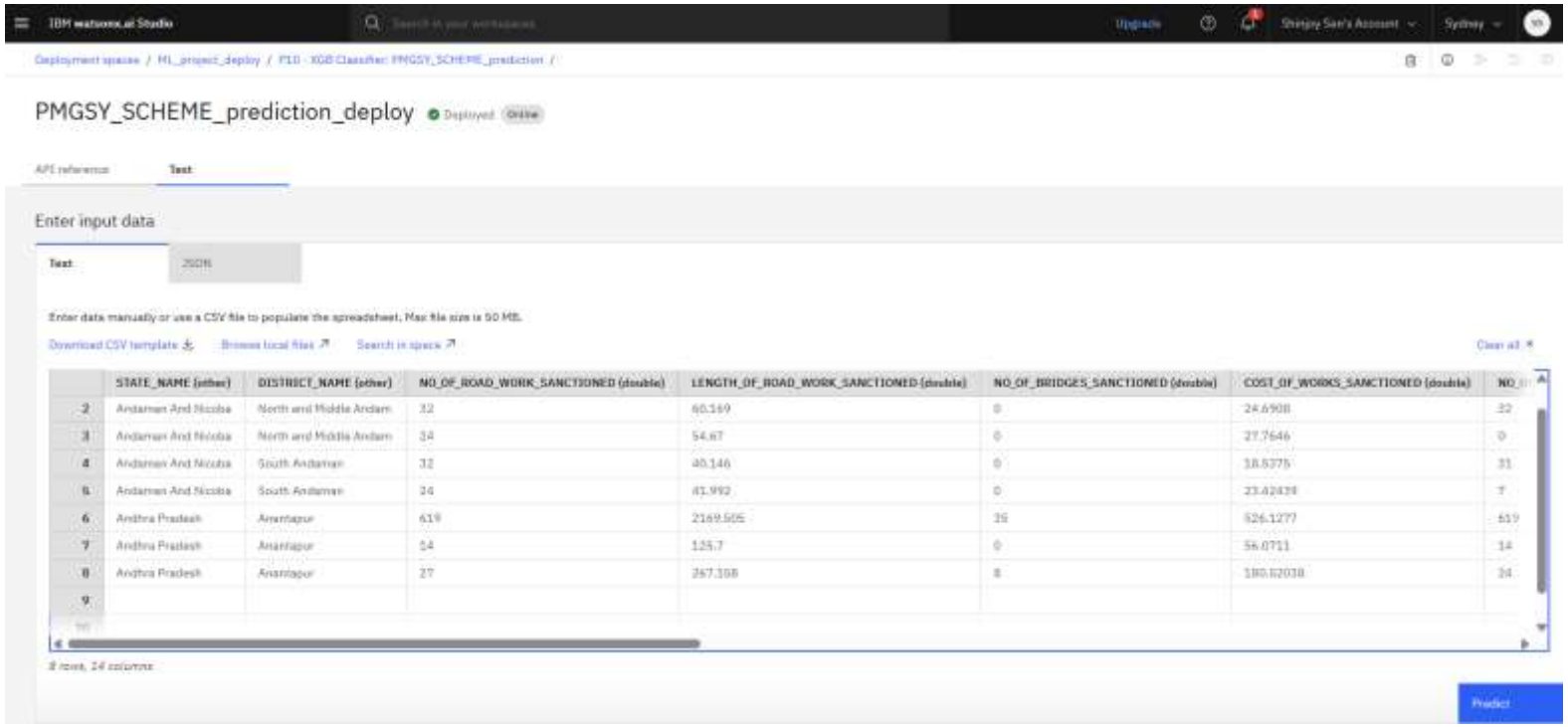
RESULT



Now we deploy our model in the deployment space in IBM cloud.

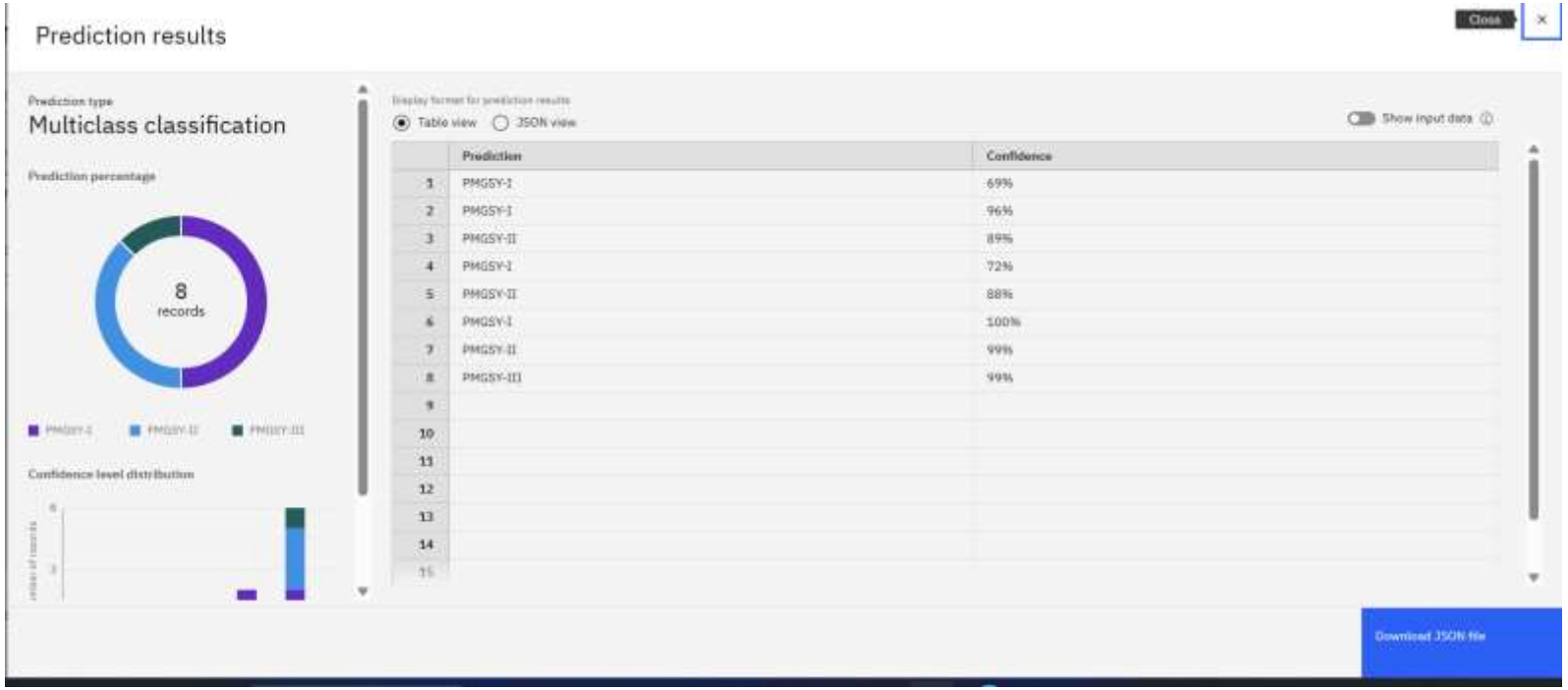
edunet foundation

RESULT



Provide the input features to predict the output.

RESULT



Based on the given input feature the output column is predicted with the mentioned accuracy(confidence).

CONCLUSION

The proposed solution effectively addresses the challenge of classifying infrastructure projects into their correct PMGSY_SCHEME using a machine learning-based approach on IBM Cloud. Historical data, including the number of road works, bridges, and financial expenditures, was collected and preprocessed to remove inconsistencies and enhance data quality. Feature engineering and hyperparameter tuning were applied to build robust models using Random Forest and XGBoost classifiers. Among the generated pipelines, the XGBoost-based model achieved the highest accuracy of 92.4%. This model was deployed using IBM Watson Machine Learning, enabling real-time predictions based on project characteristics. The solution demonstrates high effectiveness in automating the classification process, offering scalability, adaptability, and integration potential for decision-making and policy implementation in infrastructure planning.

FUTURE SCOPE

The proposed machine learning solution holds significant potential for future expansion and impact. As more granular and diverse datasets become available—such as geographic information, terrain type, contractor performance history, and weather conditions—the model can be further refined to improve accuracy and contextual relevance. Integration with GIS (Geographic Information Systems) and real-time IoT sensors from ongoing projects can enable predictive insights and anomaly detection during execution phases. Additionally, automating model retraining with new data through MLOps pipelines on IBM Cloud will ensure sustained accuracy and adaptability. The solution can also be extended to predict project timelines, risk levels, or budget overruns, offering a comprehensive decision-support tool for government agencies. Ultimately, the model can aid in smarter resource allocation, improved transparency, and data-driven policymaking for rural infrastructure development.

REFERENCES

- Dataset used: AI Kosh dataset link – https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_y ojna_pmgsy.html

- Multiple ML model used: : xgboost, sklearn.ensemble.RandomForestClassifier etc.

IBM CERTIFICATIONS

- Screenshot/ credly certificate(getting started with AI)

IBM CERTIFICATIONS

- Screenshot/ credly certificate(Journey to Cloud)

edunet foundation

IBM CERTIFICATIONS

- Screenshot/ credly certificate(RAG Lab)

edunet foundation

THANK YOU

edunet foundation