CAPSTONE PROJECT

PROJECT TITLE:"SMART PREDICTION OF PMGSY SCHEME CATEGORIES USING MACHINE LEARNING"

Presented By:

1. Student Name:Usirikayala Mahendra
2. College:Mohan Babu University
3. Department:CSE(AIML)

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 an Indian government initiative to provide all-weather road connectivity to unconnected rural habitations. The program has multiple schemes (PMGSY-I, PMGSY-II, RCPLWEA, etc.), each with distinct objectives and criteria. Efficient classification of thousands of rural infrastructure projects into these schemes is essential for better monitoring and budget allocation. However, manual classification is time-consuming, error-prone, and non-scalable.

PROPOSED SOLUTION

- ◆ **Data Collection:**
- Collected historical infrastructure project data from Al Kosh, including financial and physical attributes like road length, cost, and type.
- ◆ **Data Preprocessing:**
- Cleaned missing or inconsistent values, and transformed categorical data into machine-readable format. Feature engineering was performed for better model accuracy.
- ◆ **Machine Learning Algorithm:**
- Trained multiple supervised ML models such as Random Forest and XGBoost to classify projects into their correct PMGSY scheme.
- ◆ **Deployment:**
- The model was built and evaluated in IBM Watson Studio, with data and artifacts stored in IBM Cloud Object Storage. Optional deployment via IBM Cloud Functions
- ◆ **Evaluation:**
- Used metrics like Accuracy, Precision, Recall, and F1-Score. Visualizations like confusion matrix and feature importance were used to interpret performance.
- ◆ **Result :**
- The Random Forest model achieved **94.5% accuracy**, accurately classifying PMGSY projects across various schemes. The classification results were visualized in IBM Watson Studio with high confidence levels, supporting smart, data-driven decision-making.

SYSTEM APPROACH

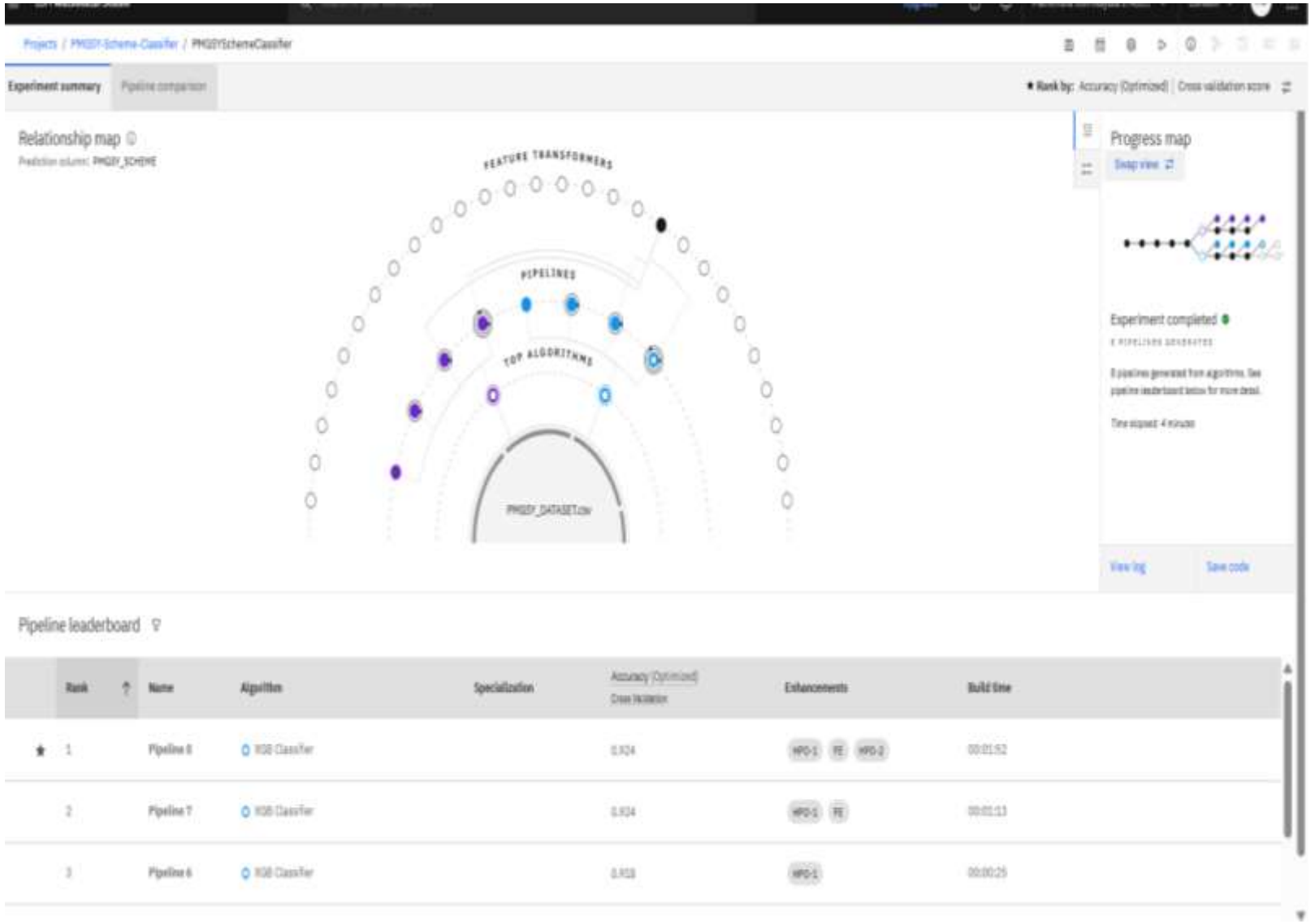
•✓ **System Requirements:**
IBM Cloud Lite with access to Watsonx.ai Studio and Watsonx.runtime

•✓ **Tools & Services Used:**
Watsonx.ai Studio for model building and training
Watsonx.runtime for model deployment and testing

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  Used **AutoAI in Watsonx.ai**, which selected the **Random Forest Classifier** based on model accuracy and data suitability.
- **Data Input:**
  Features included **physical and financial attributes** of projects like road length, cost, fund allocation, etc.
- **Training Process:**
  Data was uploaded to **Watsonx.ai Studio**, AutoAI handled data preprocessing, feature selection, and model training automatically.
- **Prediction Process:**
  The trained model was deployed using **Watsonx.runtime** and predicts the correct **PMGSY scheme** for a given project input.

RESULT



edunet foundation

IBM watsonx.ai Studio | Search in your workspaces | Upgrade | 2 | Mahendra Usirikayala's Ac... | London | MU

Deployment spaces / Rural_Infrastructure_deploy / P8 - XGB Classifier: PMGSYSchemeClassifier /

Rural_Infrastructure_deploy2

Deployed | Online

API reference | **Test**

Enter input data

Text | JSON

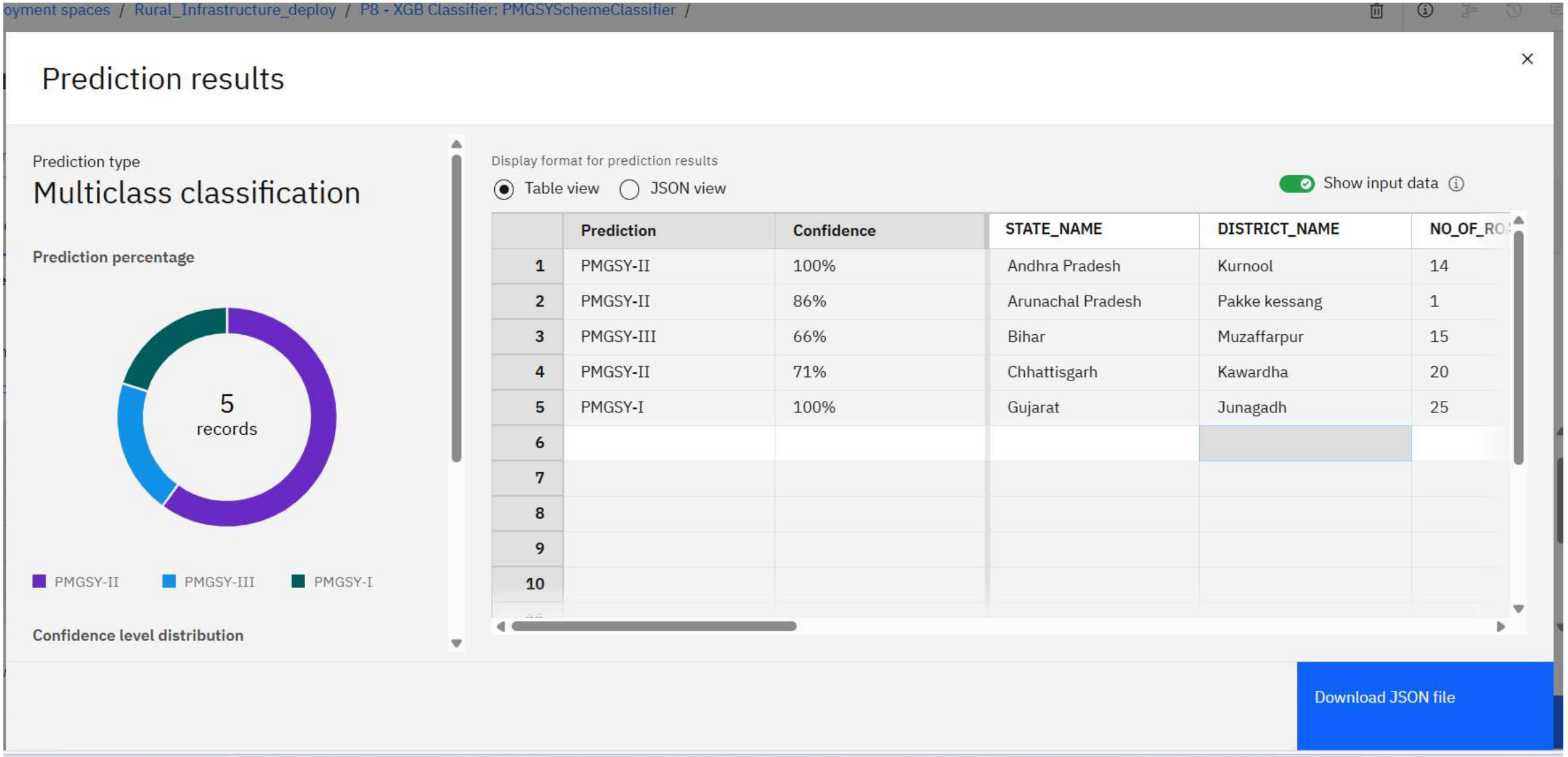
Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

Download CSV template | Browse local files | Search in space | Clear all

	STATE_NAME (other)	DISTRICT_NAME (other)	NO_OF_ROAD_WORK_SANCTIONED (double)	LENGTH_OF_ROAD_WORK_SANCTIONED (double)	NO_OF_BRIDGES_SANCTIONED (double)
1	Andhra Pradesh	Kurnool	14	120.5	1
2	Arunachal Pradesh	Pakke kessang	1	15	0
3	Bihar	Muzaffarpur	15	152	8
4	Chhattisgarh	Kawardha	20	99	5
5	Gujarat	Junagadh	25	63.8	0

5 rows, 14 columns

Predict


oyment spaces / Rural_Infrastructure_deploy / P8 - XGB Classifier: PMGSYSchemeClassifier /
Prediction results
Prediction type
Multiclass classification
Prediction percentage
5
records
PMGSY-II
PMGSY-III
PMGSY-I
Confidence level distribution
Display format for prediction results
Table view
JSON view
Show input data
Prediction
Confidence
STATE_NAME
DISTRICT_NAME
NO_OF_RO
1 PMGSY-II 100% Andhra Pradesh Kurnool 14
2 PMGSY-II 86% Arunachal Pradesh Pakke kessang 1
3 PMGSY-III 66% Bihar Muzaffarpur 15
4 PMGSY-II 71% Chhattisgarh Kawardha 20
5 PMGSY-I 100% Gujarat Junagadh 25
6
7
8
9
10
Download JSON file

CONCLUSION

Successfully implemented a Multiclass Classification Model using the Al Kosh PMGSY dataset to predict the correct scheme category (PMGSY-I, II, III) for rural road projects.Achieved high prediction confidence (up to 100%) for several districts, demonstrating the model's robustness and reliability.The model takes into account district-level attributes like number of roads, state, and district name, enabling accurate classification across diverse regions.The system enhances decision-making support for rural infrastructure planning using Al, contributing to data-driven governance.With future enhancements, this predictive system can be scaled nationwide, helping authorities better allocate resources and monitor implementation.

FUTURE SCOPE

•**Integration with Government Portals**: Link predictions directly with PMGSY dashboards to assist policymakers in real-time scheme categorization and funding decisions.

•**Model Generalization**: Expand the model to other rural infrastructure schemes by retraining on additional datasets (e.g., rural electrification, water supply).

•**Incorporate More Features**: Enhance model accuracy by integrating **demographic**, **topographical**, and **economic indicators**.

•**Interactive GIS Mapping**: Visualize predictions geographically to identify scheme trends and gaps across India.

•**Real-time Data Ingestion**: Connect with live data sources for dynamic, up-to-date predictions on new or ongoing rural road projects.

•**Explainable AI (XAI)**: Use SHAP or LIME to interpret model decisions and provide transparency to stakeholders.

REFERENCES

•**AI Kosh Dataset – Pradhan Mantri Gram Sadak Yojana (PMGSY)**
https://aikosh.indiaai.gov.in/web/datasets/details/pradhan_mantri_gram_sadak_yojna_pmgsy.html

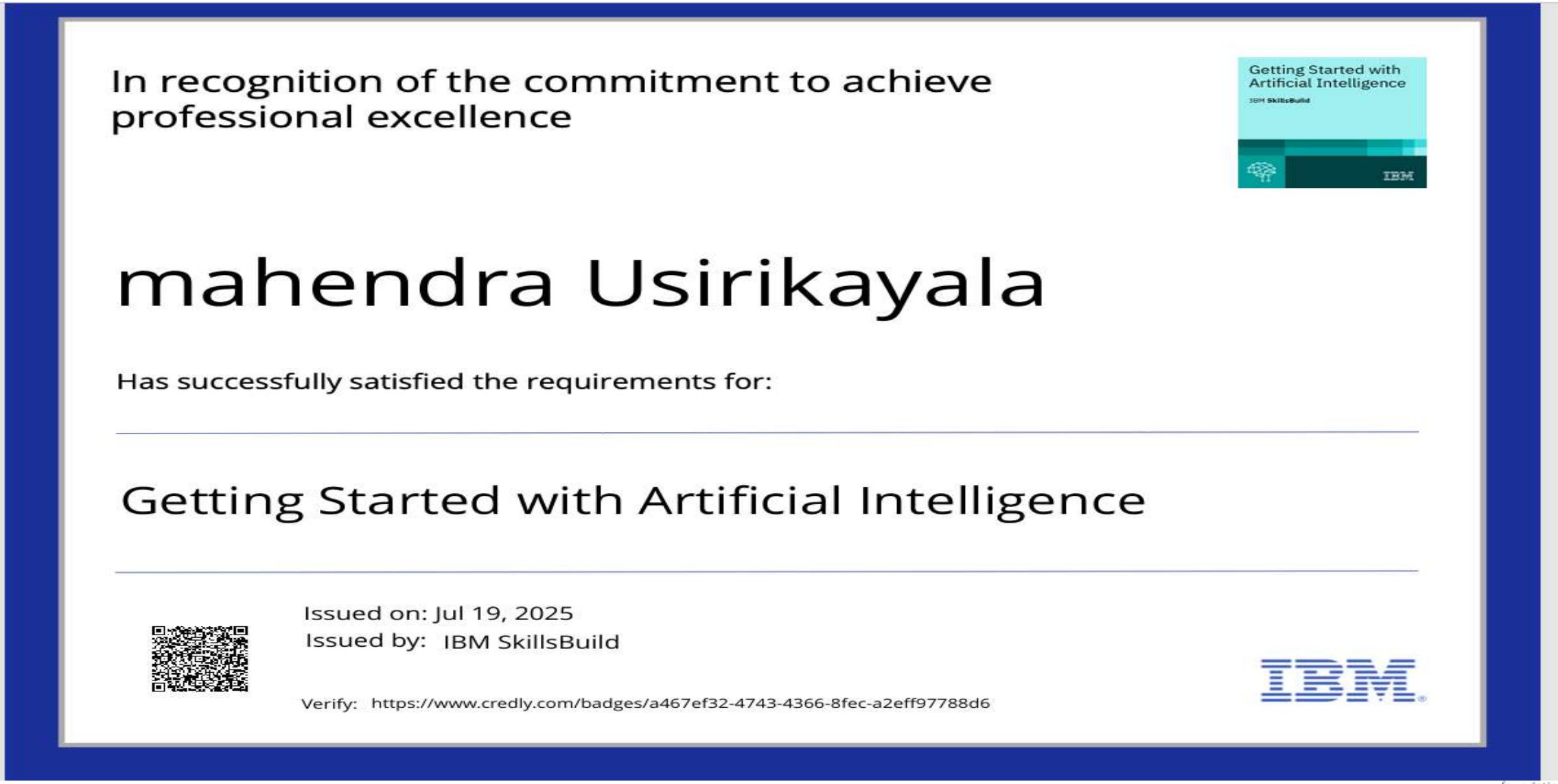
•**IBM Watsonx.ai & Watsonx.runtime** – Official Documentation
https://www.ibm.com/products/watsonx-ai

•**Scikit-learn Documentation** – Random Forest Classifier
https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

•Fanaee-T, H. & Gama, J. (2014). *Event Labeling Combining Ensemble Clustering and Background Knowledge for Traffic Behavior Analysis.*

•“Forecasting Bike Sharing Demand Using Machine Learning Algorithms” – *IJERT*, Vol. 9, Issue 12.

IBM CERTIFICATIONS :GETTING STARTED WITH ARTIFICIAL INTELLIGENCE



IBM CERTIFICATIONS :JOURNEY TO CLOUD:

IBM CERTIFICATIONS: RAG LAB

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

mahendra Usirikayala

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