CAPSTONE PROJECT

ENHANCING NSAP DELIVERY WITH AI: PREDICTING PENSION SCHEME ELIGIBILITY USING IBM CLOUD SERVICES

Presented By:

NAME :YAGATI DILEEP KUMAR

COLLEGE NAME :VELAGAPUDI RAMAKRISHNA SIDDHARTHA ENGINEERING COLLEGE

BRANCH :CSE[AI&ML]

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 National Social Assistance Program (NSAP) is a flagship social security and welfare program by the Government of India. It aims to provide financial assistance to the elderly, widows, and persons with disabilities belonging to below-poverty-line (BPL) households. The program consists of several sub-schemes, each with specific eligibility criteria. Manually verifying applications and assigning the correct scheme can be a time consuming and error-prone process. Delays or incorrect allocation can prevent deserving individuals from receiving timely financial aid. Your task is to design, build, and evaluate a multi-class classification model that can accurately predict the most appropriate NSAP scheme for an applicant based on their demographic and socio-economic data. The goal is to create a reliable tool that could assist government agencies in quickly and accurately categorizing applicants, ensuring that benefits are delivered to the right people efficiently.

PROPOSED SOLUTION

- The proposed system aims to solve the challenge of accurately assigning the most suitable NSAP sub-scheme to applicants based on their demographic and socio-economic information. This involves using machine learning techniques to automate classification and reduce manual errors. The solution will consist of the following components:
- **Data Collection:**
  - Collect applicant data including age, gender, marital status, disability status, income level, and household type (BPL/APL).
  - Integrate additional data if available (e.g., prior benefit status, breadwinner status, dependents).
- **Data Preprocessing:**
  - Clean and preprocess the data to handle missing values, inconsistencies, and noise.
  - Perform feature engineering to derive meaningful inputs such as disability eligibility, age group category, and widowhood status.
- **Machine Learning Algorithm:**
  - Implement a multi-class classification algorithm such as Random Forest, XGBoost, or Neural Networks to predict the most appropriate NSAP scheme.
  - Consider using class balancing techniques if class distribution is skewed.
  - Use cross-validation to ensure the robustness of the model.
- **Evaluation:**
  - Evaluate the model using metrics like Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.
  - Continuously monitor model performance and fine-tune it with new data or scheme changes.
- **Result :**
  - An intelligent, automated decision-support system that helps government agencies allocate NSAP benefits faster and more accurately, ensuring that eligible citizens receive timely assistance.

SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the NSAP welfare scheme classification or beneficiary prediction system.
Here's a suggested structure for this section:

System requirements:

- **IBM Cloud (mandatory)**
  For hosting and managing cloud-native machine learning applications.

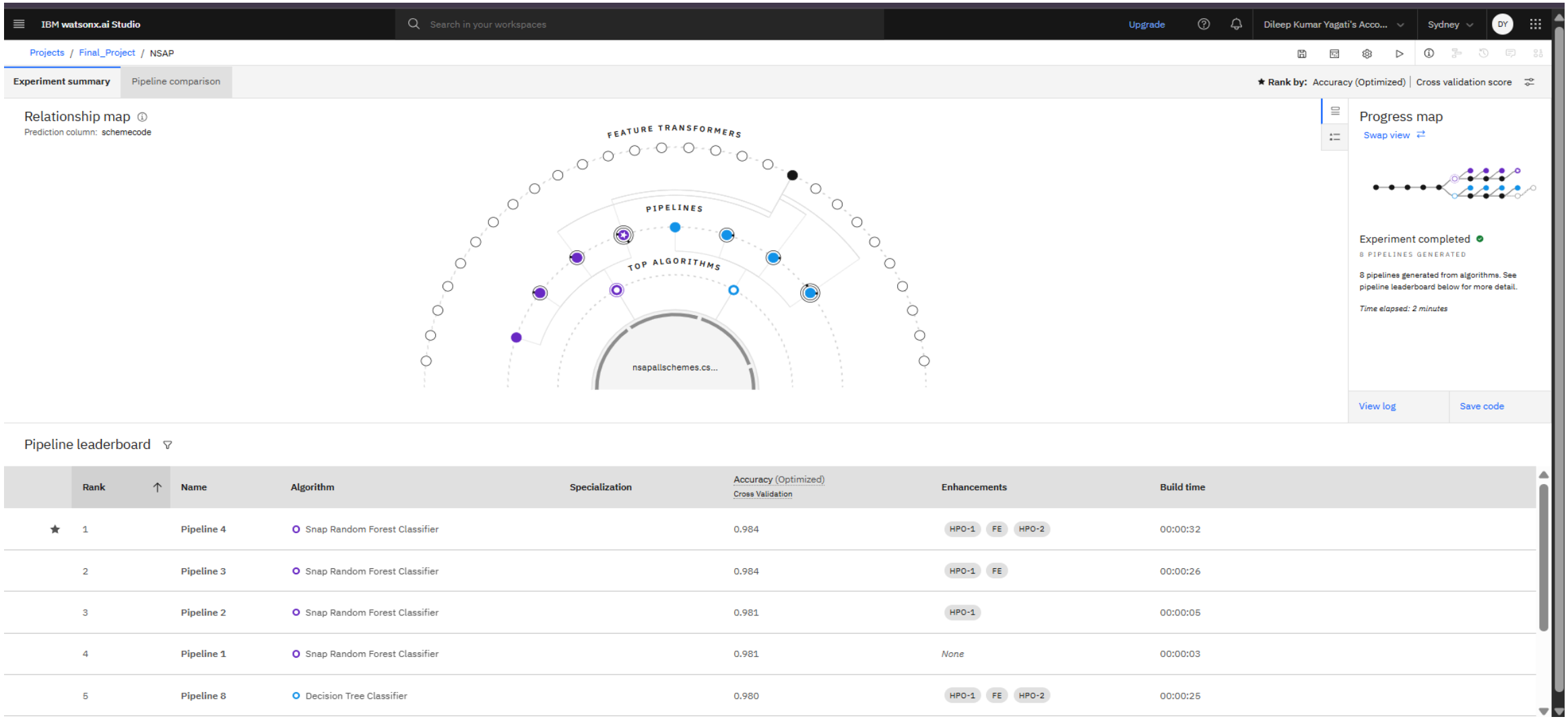
- **IBM Watson Studio**
  For model development, training, and deployment of classification or regression models.

- **IBM Cloud Object Storage**
  For storing and accessing the nsapallschemes.csv dataset during preprocessing and model training.

ALGORITHM & DEPLOYMENT

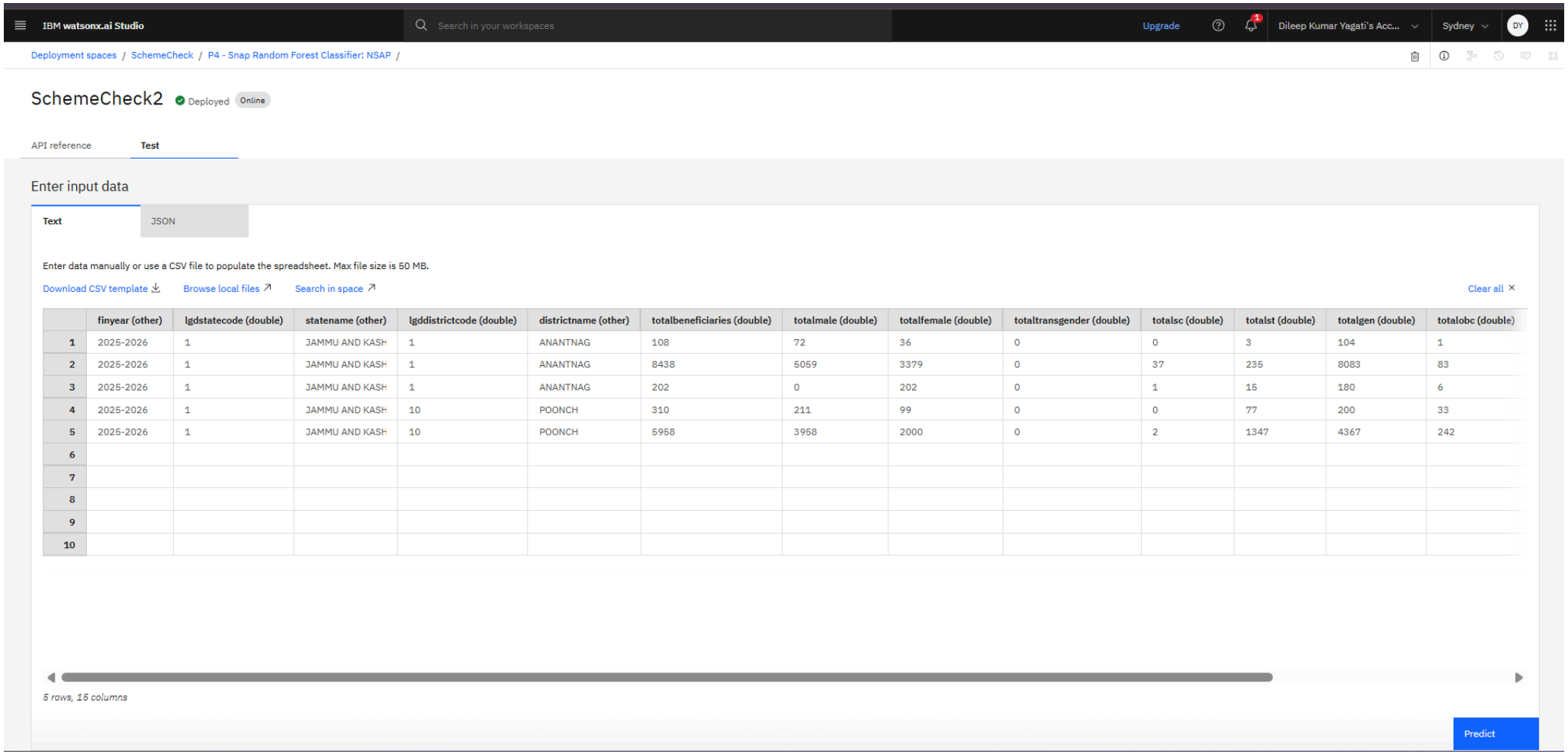
- In the Algorithm section, describe the machine learning algorithm used for predicting either the welfare scheme (schemecode. Here's an example structure for this section:
- **Algorithm Selection:**
  - Provide a brief overview of the chosen algorithm (e.g., Decision Tree, Random Forest, Logistic Regression, or XGBoost).
  - Justify the selection based on data characteristics such as categorical attributes (statename, districtname, schemecode) and numeric counts (e.g., gender-wise and caste-wise counts).
- **Data Input:**
  - Specify the input features used, including : statename , districtname , schemecode and Demographic features like totalmale , totalfemale, totaltransgender and Social category counts: totalsc, totalst, totalgen, totalobc and Digital accessibility features: totalaadhaar , totalmobilenumber
- **Training Process:**
  - Explain how the algorithm is trained using the available dataset.Mention preprocessing steps such as:
    - One-hot encoding for categorical variables
    - Standardization or normalization for numerical values
    - Train-test split or cross-validation
    - Highlight any tuning done (e.g., grid search for optimal parameters).
- **Prediction Process**:
  - Describe how the trained model makes predictions, e.g.: Predicting the likely schemecode based on region and demographics (classification).

RESULT



	Rank	Name	Algorithm	Specialization	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1	Pipeline 4	Snap Random Forest Classifier		0.984	HPO-1 FE HPO-2	00:00:32
	2	Pipeline 3	Snap Random Forest Classifier		0.984	HPO-1 FE	00:00:26
	3	Pipeline 2	Snap Random Forest Classifier		0.981	HPO-1	00:00:05
	4	Pipeline 1	Snap Random Forest Classifier		0.981	*None*	00:00:03
	5	Pipeline 8	Decision Tree Classifier		0.980	HPO-1 FE HPO-2	00:00:25

RESULT



	finyear (other)	lgdstatecode (double)	statename (other)	lgddistrictcode (double)	districtname (other)	totalbeneficiaries (double)	totalmale (double)	totalfemale (double)	totaltransgender (double)	totalsc (double)	totalst (double)	totalgen (double)	totalobc (double)
1	2025-2026	1	JAMMU AND KASH	1	ANANTNAG	108	72	36	0	0	3	104	1
2	2025-2026	1	JAMMU AND KASH	1	ANANTNAG	8438	5059	3379	0	37	235	8083	83
3	2025-2026	1	JAMMU AND KASH	1	ANANTNAG	202	0	202	0	1	15	180	6
4	2025-2026	1	JAMMU AND KASH	10	POONCH	310	211	99	0	0	77	200	33
5	2025-2026	1	JAMMU AND KASH	10	POONCH	5958	3958	2000	0	2	1347	4367	242
6													
7													
8													
9													
10													

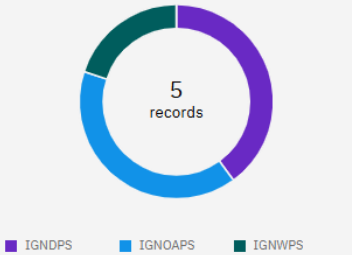
RESULT

Prediction results

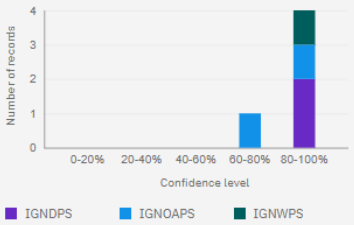
Prediction type

Multiclass classification

Prediction percentage



Confidence level distribution



Display format for prediction results

(•) Table view ( ) JSON view

Show input data

	Prediction	Confidence	finyear	lgdstatecode	statename	lgddistrictcode	districtname	totalbeneficiaries	totalmale
1	IGNDPS	100%	2025-2026	1	JAMMU AND KASHMIR	1	ANANTNAG	108	72
2	IGNOAPS	70%	2025-2026	1	JAMMU AND KASHMIR	1	ANANTNAG	8438	5059
3	IGNWPS	100%	2025-2026	1	JAMMU AND KASHMIR	1	ANANTNAG	202	0
4	IGNDPS	100%	2025-2026	1	JAMMU AND KASHMIR	10	POONCH	310	211
5	IGNOAPS	100%	2025-2026	1	JAMMU AND KASHMIR	10	POONCH	5958	3958
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									

Download JSON file

edunet foundation

CONCLUSION

- Key Outcomes:
  - . Successfully built a machine learning model to predict the NSAP welfare scheme (schemecode) using demographic and regional data.
  - Demonstrated that features like state, district, gender split, and social category have strong predictive power.
  - The system enhances the transparency, efficiency, and accuracy of beneficiary mapping across schemes
- Impact:
  - Can assist government agencies in automated scheme classification, resource allocation, and policy evaluation.
  - Reduces manual errors and speeds up beneficiary identification and approval processes.
- Conclusion:
  - The project proves the viability of AI-driven solutions in streamlining public welfare programs, ensuring timely and targeted assistance to vulnerable groups.

FUTURE SCOPE

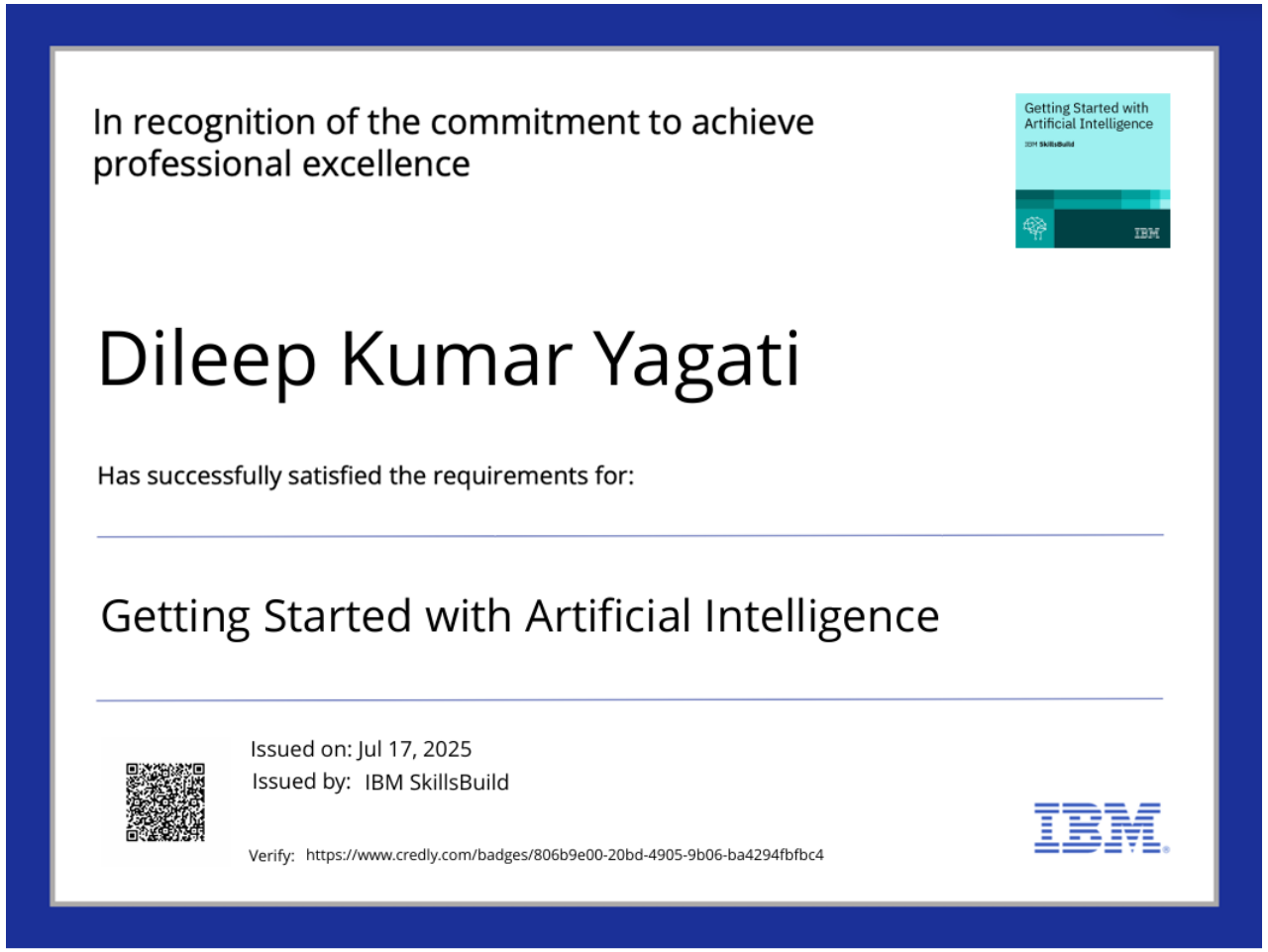
- Data Enhancements:
  - Integrate additional socio-economic indicators, health records, and real-time data from government APIs.
  - Expand dataset to include longitudinal data for trend analysis and forecasting
- Model Improvements:
  - Use advanced algorithms (e.g., XGBoost, LightGBM, or deep learning) for better classification accuracy.
  - Apply explainable AI for transparent and accountable decision-making.
- Geographic & Technological Expansion:
  - Scale the system to cover all Indian districts and union territories.
  - Enable mobile access and use edge computing for local-level deployment.
  - Explore federated learning for secure, decentralized model training across states.
- Vision:
  - Build a national AI-powered support system to optimize social welfare delivery with real-time monitoring and data-driven governance.

REFERENCES

- Dataset Source
- AI Kosh – District-wise Pension Data under the NSAP
  Ministry of Rural Development, Government of India
  https://aikosh.indiaai.gov.in/web/datasets/details/district_wise_pension_data_under_the_national_social_assistance_programme_nsap_1.html
  ➤ Used as the primary dataset for analyzing and predicting welfare scheme distribution.
- Research Papers

1. Kadari, A., & Valaboju, R. (2025)
   Leveraging Artificial Intelligence and Machine Learning to Reform India's Social Protection Framework
   ➤ Demonstrates how ML can improve accuracy in identifying eligible beneficiaries and reduce inclusion/exclusion errors in welfare systems.
2. Sansone, D., & Zhu, L. (2020)
   Using Machine Learning to Create an Early Warning System for Welfare Recipients
   ➤ Explores predictive modeling to support proactive welfare delivery based on demographic and economic indicators.

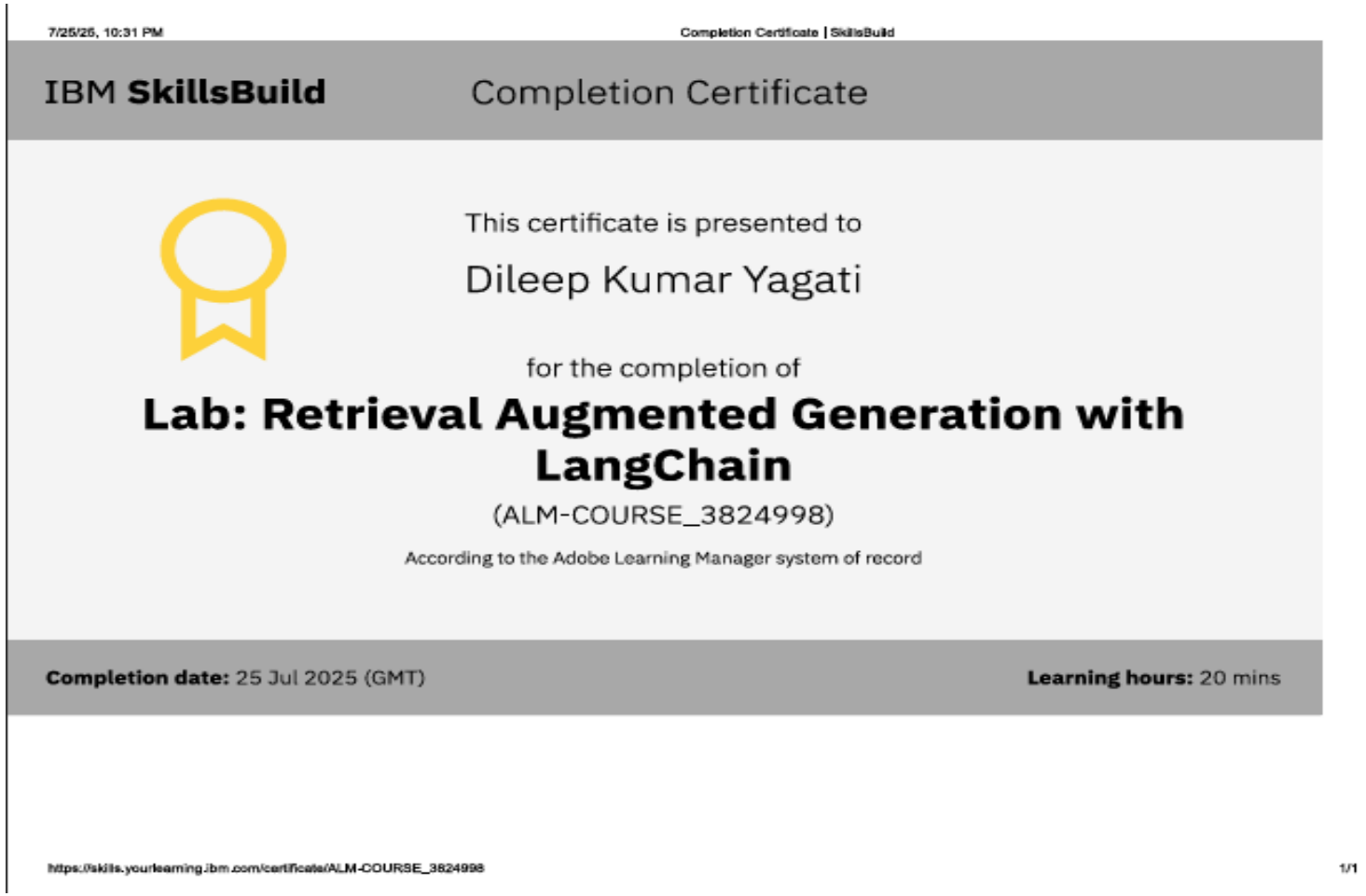
IBM CERTIFICATION : “GETTING STARTED WITH ARTIFICIAL INTELLIGENCE “



IBM CERTIFICATION : “JOURNEY TO CLOUD : ENVISIONING YOUR SOLUTION “

edunet foundation

IBM CERTIFICATION : “RAG WITH LANGCHAIN LAB”



7/25/25, 10:31 PM Completion Certificate | SkillsBuild

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Dileep Kumar Yagati

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

edunet foundation

THANK YOU

edunet foundation