CAPSTONE PROJECT

AUTOMATED ELIGIBILITY PREDICTION FOR NSAP SCHEMES USING MACHINE LEARNING

Presented By:

1. Siva Prakash S
- University College of Engineering(BIT Campus), Tiruchirappalli
- B. Tech. Information Technology(IT)

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 welfare initiative that provides financial support to the elderly, widows, and persons with disabilities from BPL households. However, manually identifying eligibility for different schemes under NSAP is time-consuming, error-prone, and often leads to incorrect or delayed distribution of benefits. This project addresses the need to automate the eligibility classification process to ensure efficient, accurate, and timely allocation of financial aid.

PROPOSED SOLUTION

Proposed Solution:

- The proposed system aims to address the challenge of accurately and efficiently predicting the most appropriate NSAP scheme for an applicant using demographic and socio-economic data. This solution leverages machine learning techniques to classify applicants into eligible scheme categories and eliminates the need for manual verification.

Data Collection:

- Utilize the AI Kosh dataset containing district-wise records of NSAP beneficiaries, including total beneficiaries, gender-wise counts, caste categories, Aadhaar coverage, and mobile number availability.
- Gather structured data that reflects socio-economic status and demographic characteristics to serve as features for classification.

Data Preprocessing:

- Clean the dataset by handling missing values and inconsistencies.
- Perform normalization or scaling where necessary.
- Conduct feature engineering to derive relevant features (e.g., SC/ST/OBC proportions, gender ratio) that may influence scheme eligibility.

Machine Learning Algorithm:

- Implement a multi-class classification model, where the target variable is the schemecode (e.g., IGNOAPS, IGNDPS, IGNWPS).
- Apply hyperparameter optimization and cross-validation to enhance model accuracy and generalization.

Deployment:

- Deploy the best-performing model as a REST API using IBM Watsonx.ai and IBM Cloud Lite.
- Allow real-time or batch processing of new applicant data through JSON or CSV input formats.
- Enable government agencies to integrate this model for seamless eligibility verification.
- Visual dashboards and predictions deployed on IBM Cloud for practical use and demonstration.

Evaluation:

- Evaluate the model using cross-validation accuracy, confusion matrix, and prediction confidence metrics.
- Monitor prediction confidence and retrain periodically with updated datasets for improved accuracy.

Result:

- Achieved a high classification accuracy of 98.4% with the Snap Random Forest Classifier.
- Successfully predicted scheme eligibility for multiple test cases with 90–100% confidence.

SYSTEM APPROACH

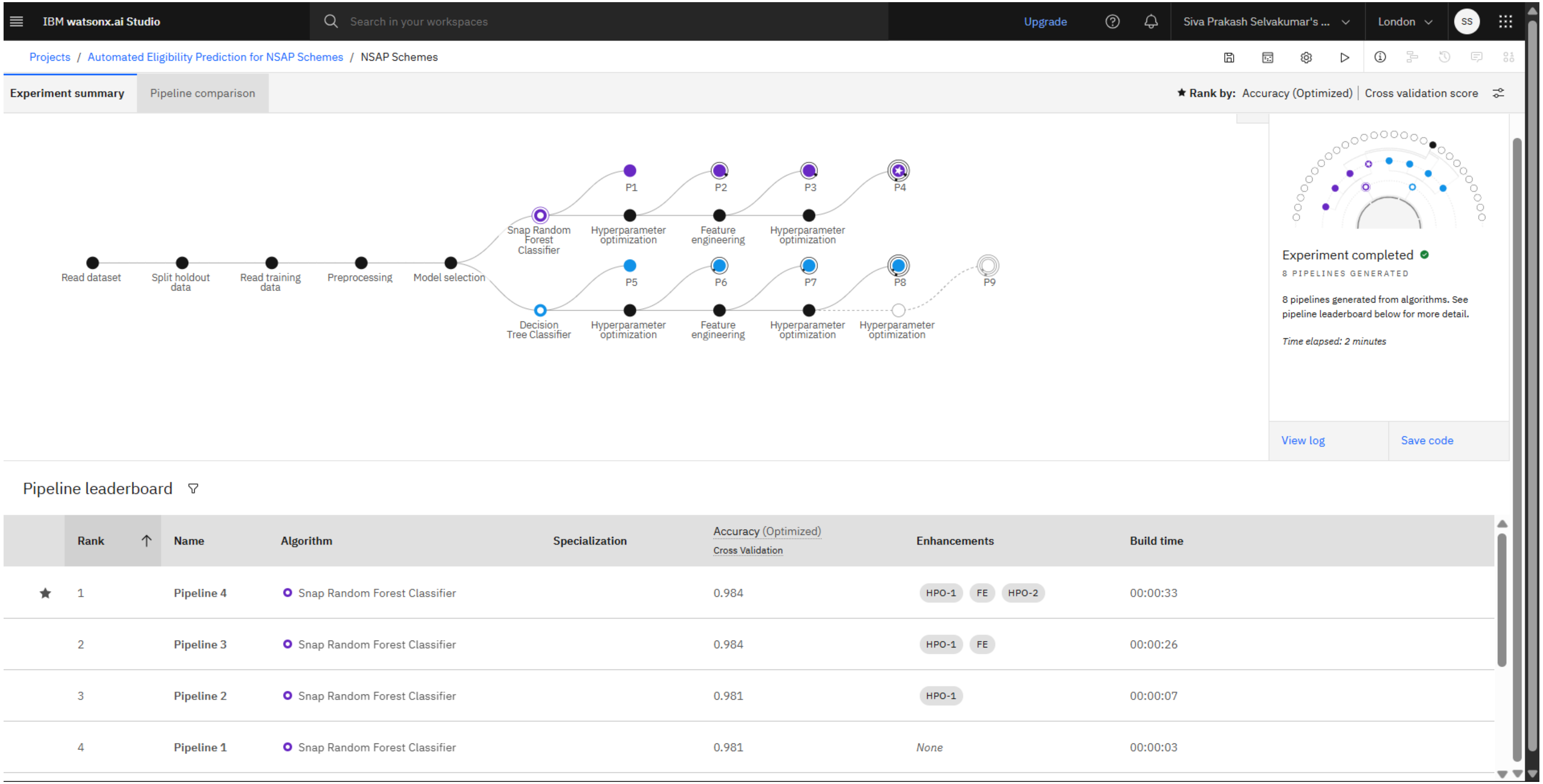
The "System Approach" section outlines the overall strategy and methodology for developing and implementing the rental bike prediction system. Here's a suggested structure for this section:

- System requirements:
  - ✓ A Laptop or System with Proper Internet Connection
  - ✓ IBM Cloud(mandatory)
  - ✓ IBM Watson studio for model development and Deployment
  - ✓ IBM Cloud object storage for Dataset Handling
  - ✓ Dataset: AI Kosh – District-wise pension data under NSAP
- Process Flow:
  - ✓Data ingestion & preprocessing
  - ✓Model training using AutoAI pipelines
  - ✓Hyperparameter optimization and model selection
  - ✓Deployment as a REST API on IBM Cloud

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
  The project uses a Snap Random Forest Classifier, automatically selected by IBM AutoAI for its high accuracy (98.4%) in multi-class classification. It’s ideal for structured tabular data and suitable for predicting NSAP scheme codes.
- **Data Input:**
  Key features include:
  totalmale, totalfemale, totaltransgender, totalsc, totalst, totalobc, totalgen, totalaadhaar, totalmobilenumber, and optionally statename, districtname, and finyear.
- **Training Process:**
  IBM AutoAI handled data preprocessing, feature engineering, and generated multiple pipelines. Cross-validation and hyperparameter tuning were used to select the most accurate model.
- **Prediction Process:**
  The final model is deployed as a REST API on IBM Cloud. It accepts new applicant data (CSV/JSON), processes the input, and returns the predicted NSAP scheme with confidence levels.

RESULT


IBM watsonx.ai Studio
Search in your workspaces
Upgrade
Siva Prakash Selvakumar's ...
London
SS
Projects / Automated Eligibility Prediction for NSAP Schemes / NSAP Schemes
Experiment summary
Pipeline comparison
★ Rank by: Accuracy (Optimized) | Cross validation score
Read dataset
Split holdout data
Read training data
Preprocessing
Model selection
Snap Random Forest Classifier
Decision Tree Classifier
Hyperparameter optimization
Feature engineering
Hyperparameter optimization
Hyperparameter optimization
Feature engineering
Hyperparameter optimization
Hyperparameter optimization
P1
P2
P3
P4
P5
P6
P7
P8
P9
View log
Save code
Pipeline leaderboard

Experiment completed ✓

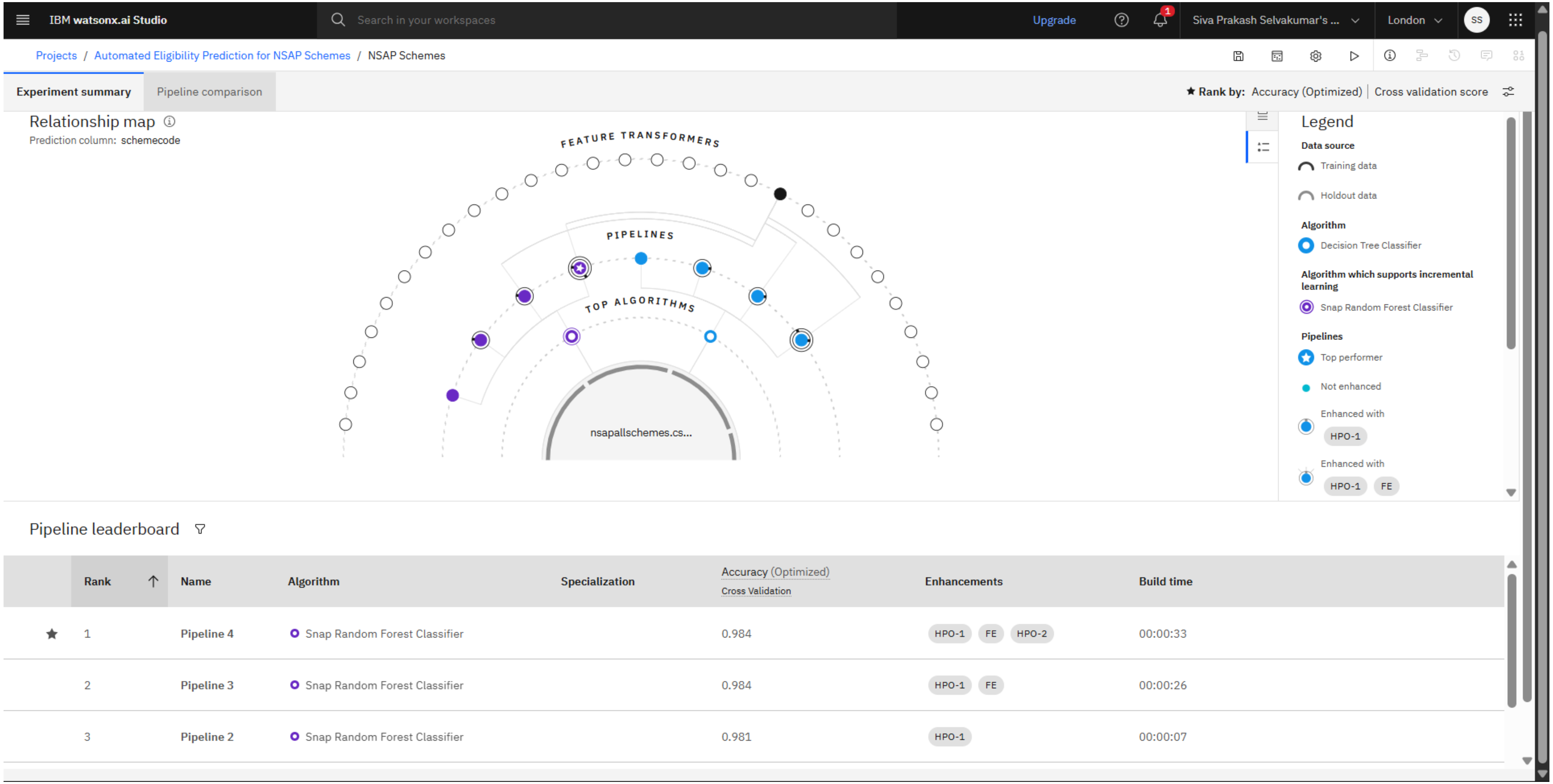
8 PIPELINES GENERATED

8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

Time elapsed: 2 minutes

	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	00:00:33
	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	00:00:07
	4	Pipeline 1	○ Snap Random Forest Classifier		0.981	*None*	00:00:03

RESULT



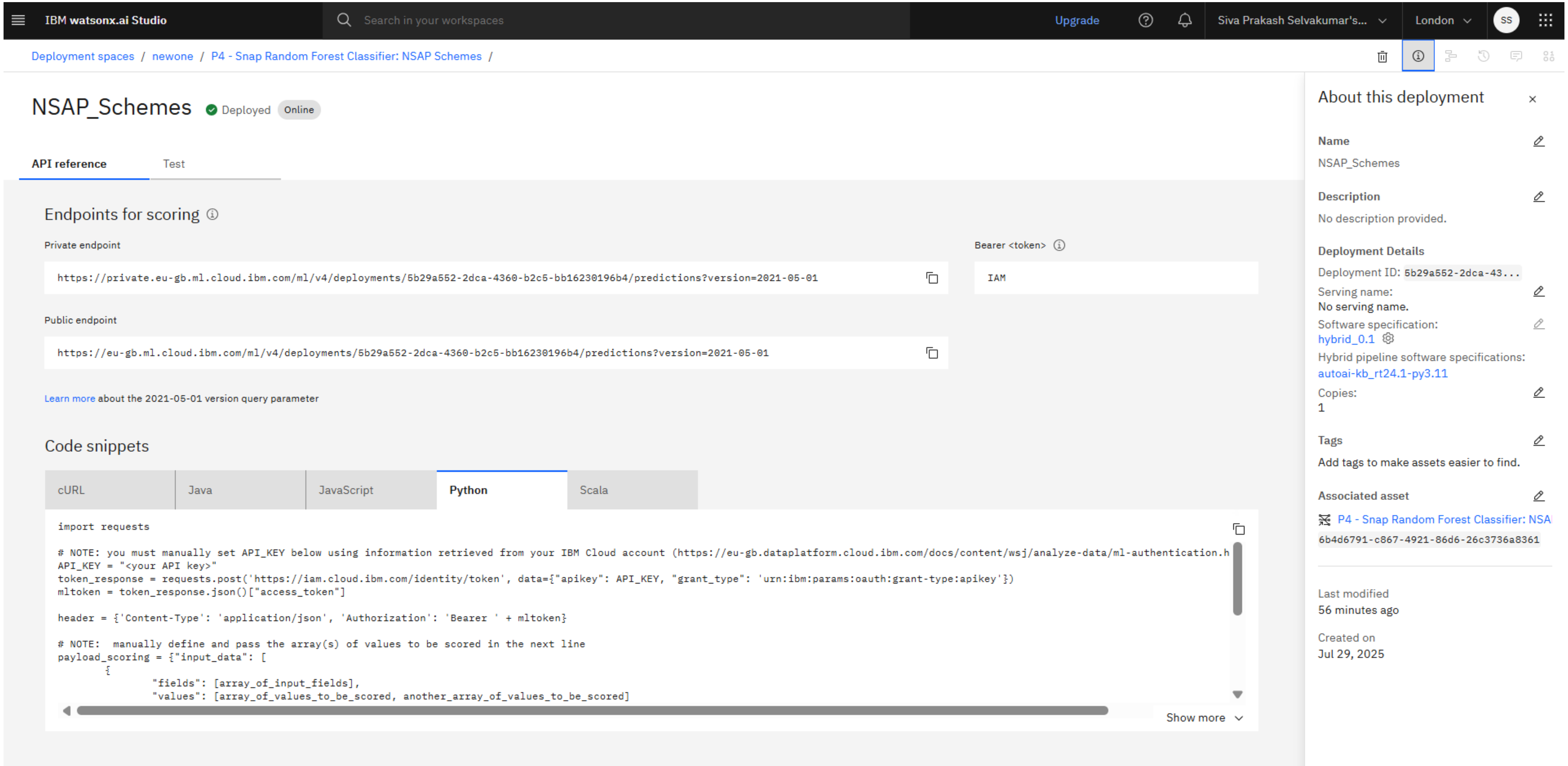
Relationship map

Prediction column: schemecode

Pipeline leaderboard

	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	00:00:33
	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	00:00:07

RESULT



Deployment spaces / newone / P4 - Snap Random Forest Classifier: NSAP Schemes /

NSAP_Schemes

Deployed Online

API reference | Test

Endpoints for scoring

Private endpoint

```
https://private.eu-gb.ml.cloud.ibm.com/ml/v4/deployments/5b29a552-2dca-4360-b2c5-bb16230196b4/predictions?version=2021-05-01
```

Bearer <token>

IAM

Public endpoint

```
https://eu-gb.ml.cloud.ibm.com/ml/v4/deployments/5b29a552-2dca-4360-b2c5-bb16230196b4/predictions?version=2021-05-01
```

Learn more about the 2021-05-01 version query parameter

Code snippets

cURL | Java | JavaScript | **Python** | Scala

```
import requests

# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account (https://eu-gb.dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/ml-authentication.h
API_KEY = "<your API key>"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey": API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]

header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}

# NOTE:  manually define and pass the array(s) of values to be scored in the next line
payload_scoring = {"input_data": [
        {
                "fields": [array_of_input_fields],
                "values": [array_of_values_to_be_scored, another_array_of_values_to_be_scored]
```

Show more

About this deployment

Name

NSAP_Schemes

Description

No description provided.

Deployment Details

Deployment ID: 5b29a552-2dca-43...

Serving name:
No serving name.

Software specification:
hybrid_0.1

Hybrid pipeline software specifications:
autoai-kb_rt24.1-py3.11

Copies:
1

Tags

Add tags to make assets easier to find.

Associated asset

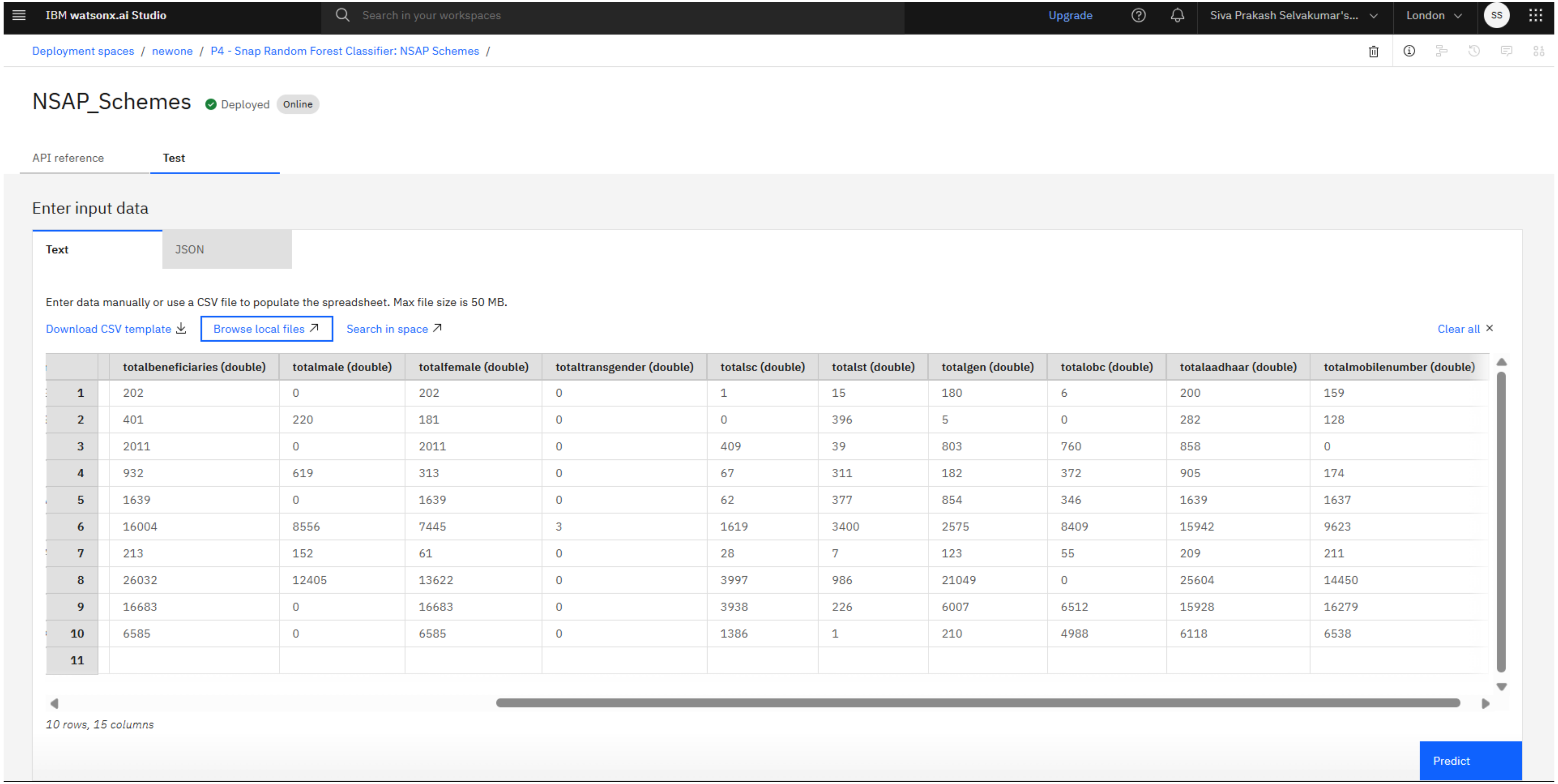
P4 - Snap Random Forest Classifier: NSA
6b4d6791-c867-4921-86d6-26c3736a8361

Last modified
56 minutes ago

Created on
Jul 29, 2025

edunet foundation

RESULT



NSAP_Schemes

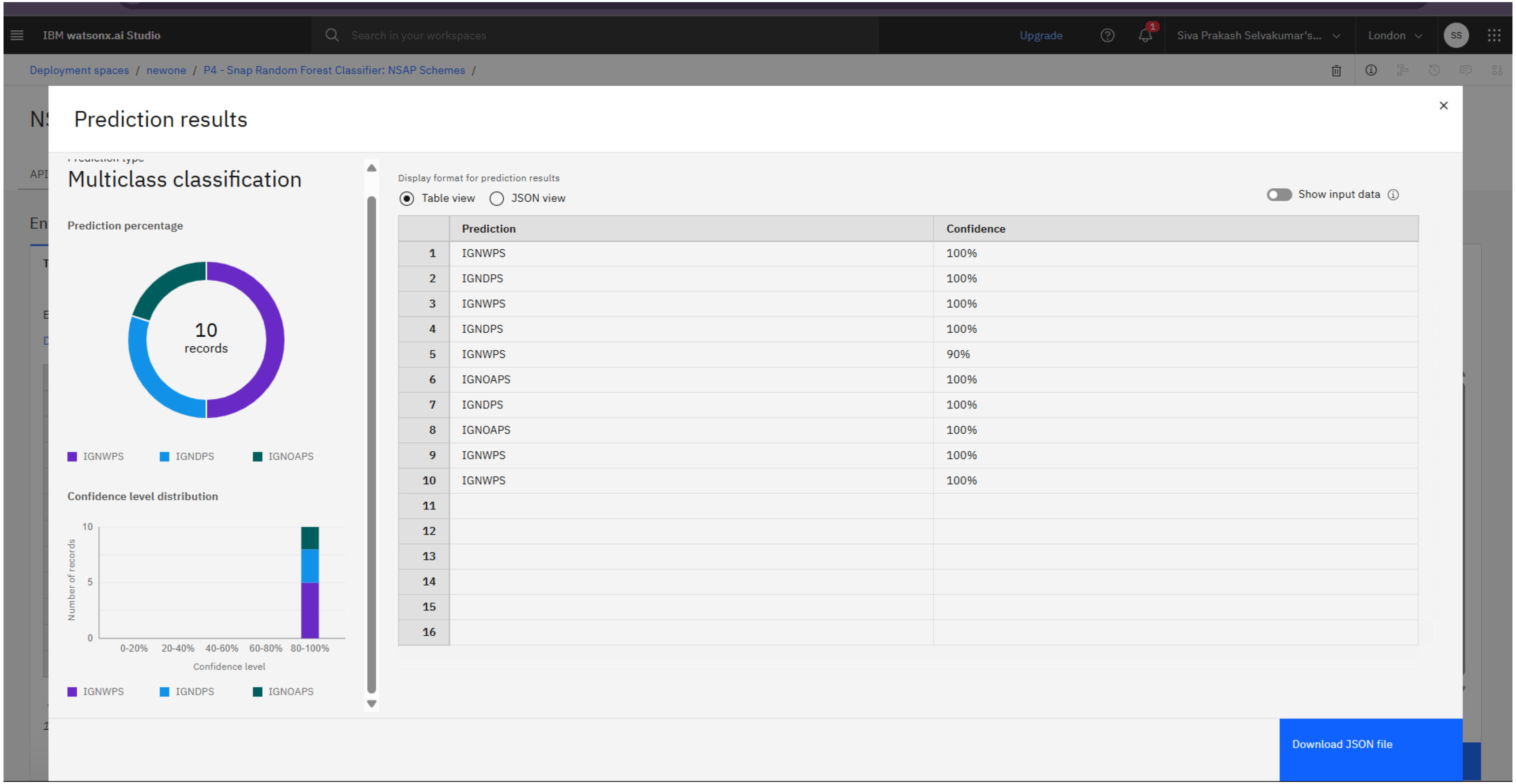
Enter input data

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

	totalbeneficiaries (double)	totalmale (double)	totalfemale (double)	totaltransgender (double)	totalsc (double)	totalst (double)	totalgen (double)	totalobc (double)	totalaadhaar (double)	totalmobilenumber (double)
1	202	0	202	0	1	15	180	6	200	159
2	401	220	181	0	0	396	5	0	282	128
3	2011	0	2011	0	409	39	803	760	858	0
4	932	619	313	0	67	311	182	372	905	174
5	1639	0	1639	0	62	377	854	346	1639	1637
6	16004	8556	7445	3	1619	3400	2575	8409	15942	9623
7	213	152	61	0	28	7	123	55	209	211
8	26032	12405	13622	0	3997	986	21049	0	25604	14450
9	16683	0	16683	0	3938	226	6007	6512	15928	16279
10	6585	0	6585	0	1386	1	210	4988	6118	6538
11										

10 rows, 15 columns

RESULT



Prediction results

Multiclass classification

Prediction percentage

Confidence level distribution

Display format for prediction results

● Table view ○ JSON view

	Prediction	Confidence
1	IGNWPS	100%
2	IGNDPS	100%
3	IGNWPS	100%
4	IGNDPS	100%
5	IGNWPS	90%
6	IGNOAPS	100%
7	IGNDPS	100%
8	IGNOAPS	100%
9	IGNWPS	100%
10	IGNWPS	100%
11		
12		
13		
14		
15		
16		

CONCLUSION

- **Model Performance:** Snap Random Forest achieved **98.4% accuracy**, proving effective for NSAP scheme prediction.
- **Deployment:** Model deployed on **IBM Watsonx.ai Studio** with a REST API for real-time, government-ready integration.
- **Challenges:**
  - Cleaning uneven demographic data
  - Handling class imbalance
  - Selecting effective features
- **Solutions:**
  - Used **AutoAI** for feature engineering and model tuning
  - Applied **cross-validation** for robust pipeline selection
- **Improvements:**
  - Add features like income, education, disability
  - Implement fairness checks
  - Schedule regular model updates
- **Impact:**
  - Speeds up and improves accuracy of benefit allocation
  - Reduces manual errors
  - Supports transparent and efficient welfare delivery

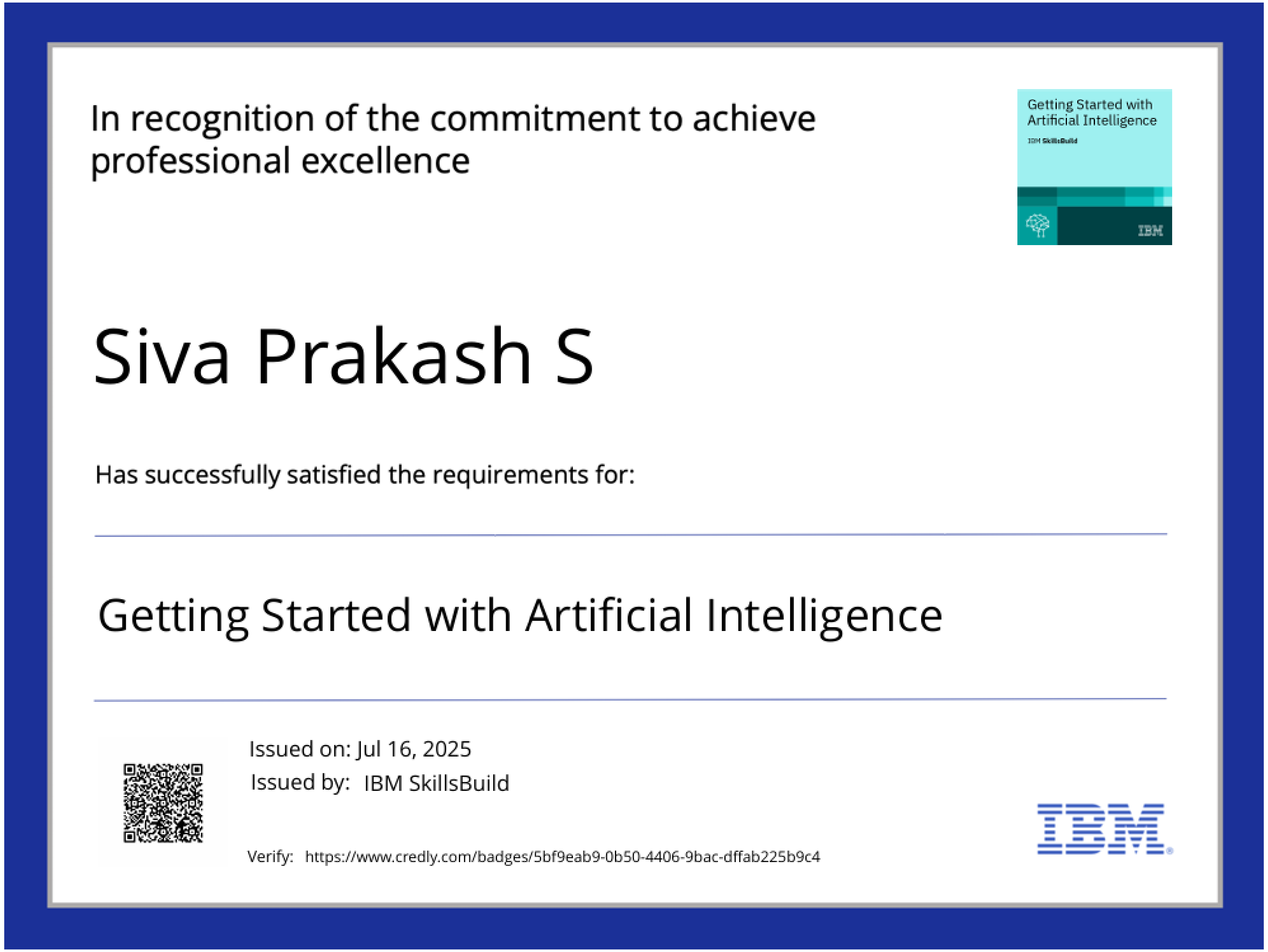
FUTURE SCOPE

- Include additional attributes like income level, education, or disability status for improved accuracy
- Integrate with live application portals for real-time eligibility checking
- Expand to other government welfare schemes (e.g., PMAY, Ujjwala Yojana)
- Incorporate fairness checks to ensure no bias in classification

REFERENCES

- Datasets & Official Resources
  - ✓ **AI Kosh Dataset:** *District-wise pension data under the National Social Assistance Programme (NSAP)*
    Source: https://aikosh.indiaai.gov.in
  - ✓ **IBM Watsonx.ai Documentation:**
    Source: https://www.ibm.com/cloud/watsonx
- Deployment & API Integration
  - ✓ IBM Cloud Docs – **Deploying and Scoring Models with Watsonx.ai**
    https://cloud.ibm.com/docs/watsonx

IBM CERTIFICATIONS


Getting Started with Artificial Intelligence
IBM SkillsBuild
IBM

In recognition of the commitment to achieve professional excellence

Siva Prakash S

Has successfully satisfied the requirements for:

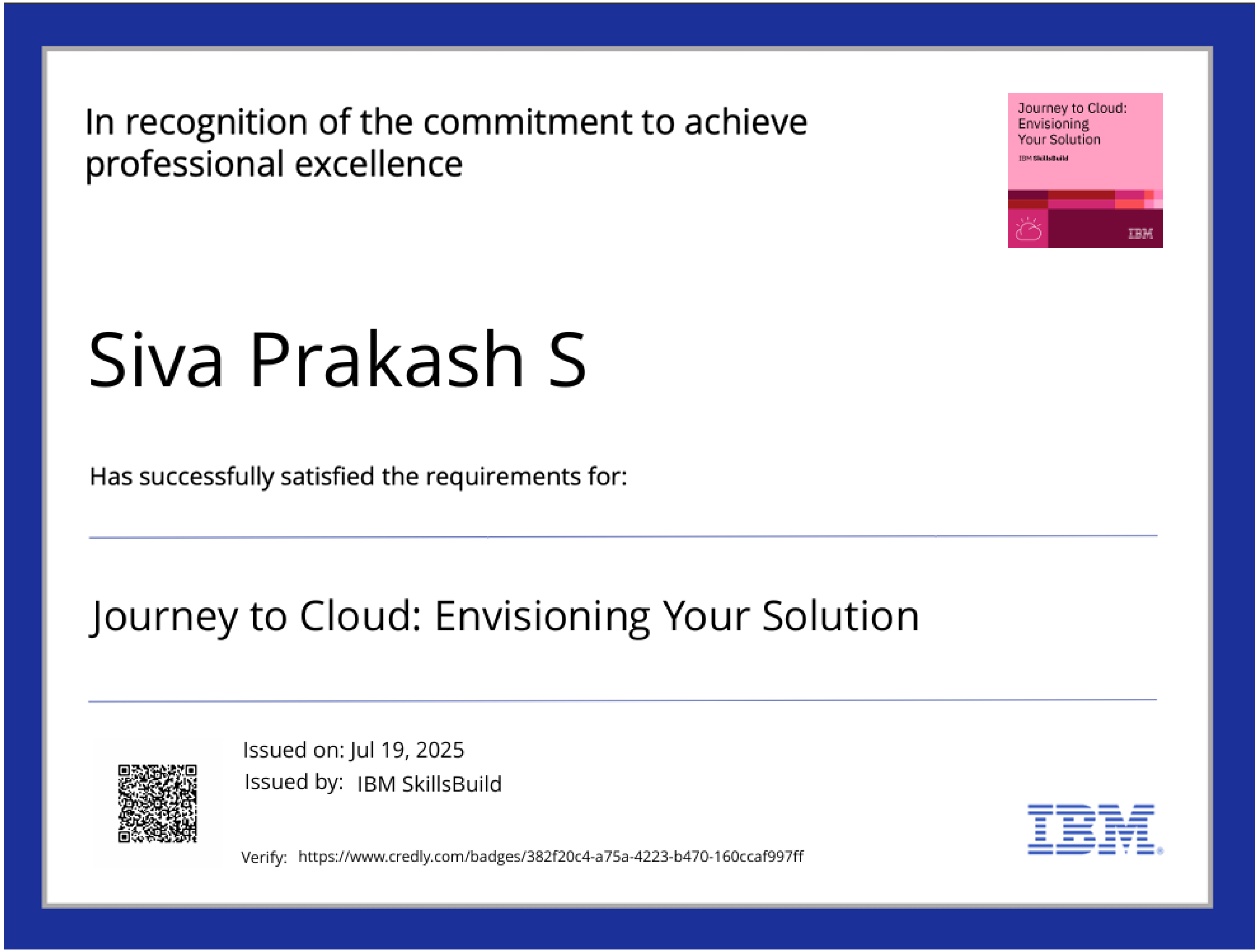
Getting Started with Artificial Intelligence

Issued on: Jul 16, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/5bf9eab9-0b50-4406-9bac-dffab225b9c4

IBM

IBM CERTIFICATIONS



Journey to Cloud: Envisioning Your Solution
IBM SkillsBuild
IBM

In recognition of the commitment to achieve professional excellence

Siva Prakash S

Has successfully satisfied the requirements for:

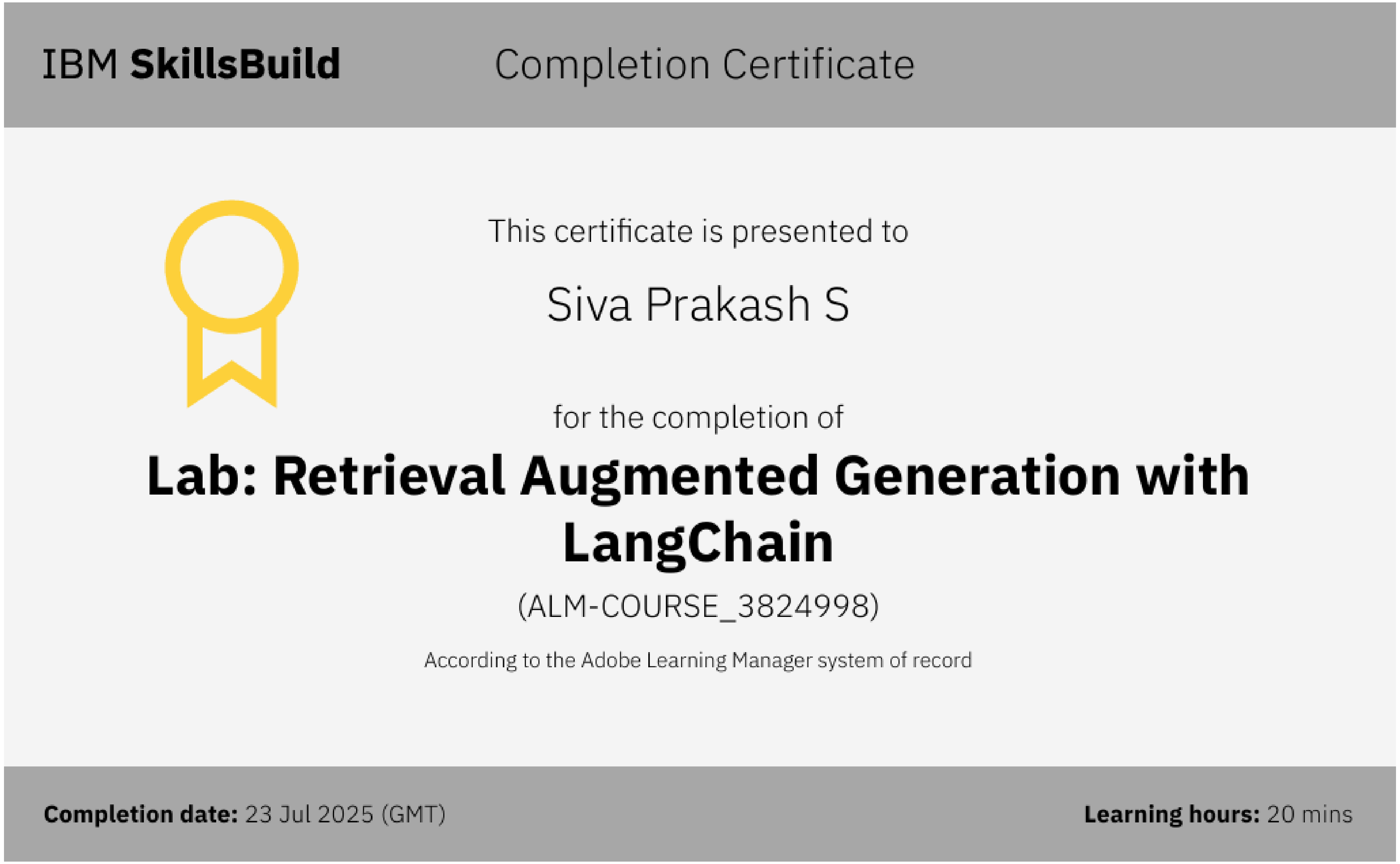
Journey to Cloud: Envisioning Your Solution

Issued on: Jul 19, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/382f20c4-a75a-4223-b470-160ccaf997ff

IBM

IBM CERTIFICATIONS



IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Siva Prakash S

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 23 Jul 2025 (GMT)

Learning hours: 20 mins

THANK YOU