

---

# **CAPSTONE PROJECT**

## **PREDICTING ELIGIBILITY FOR NSAP SCHEMES USING MACHINE LEARNING**

**Presented By:**

**STUDENT NAME: MATTAPARTHI TEJA**

**COLLEGE NAME: SWARNANDHRA COLLEGE OF ENGINEERING & TECHNOLOGY**

**DEPARTMENT : MASTER OF COMPUTER APPLICATIONS(MCA)**

# OUTLINE

- Problem Statement
- Proposed Solution
- System Development Approach
- Wow Factors & End Users
- Algorithm & Deployment
- Results
- Conclusion
- Future Scope
- Git-Hub Link
- IBM Certifications

# PROBLEM STATEMENT

## Example:

The National Social Assistance Program (NSAP) is a critical social welfare initiative in India, providing financial assistance to the elderly, widows, and persons with disabilities. The manual process of verifying applications and assigning the correct sub-scheme is a significant bottleneck. The program consists of several sub-schemes each with specific criteria, This process is time-consuming, prone to human error, and can lead to delays or incorrect allocations, preventing deserving individuals from receiving timely benefits.

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.

# PROPOSED SOLUTION

We propose building a multi-class classification machine learning model that takes in applicant data (age, income, gender, etc.) and outputs the most appropriate NSAP scheme. This model will be trained, deployed, and tested using IBM Cloud Watson Studio, enabling easy API access,

- **Data Collection:**

1. Gather Data like scheme code, total benefits, genders, personal info data in the NSAP dataset from AI Kosh to IBM Cloud Object Storage.
2. This enables secure and scalable access to the data within Watson Studio.

- **Project Creation:**

1. Create a new Watson Studio project to manage your assets and experiments.
2. This acts as the workspace where datasets, models, and deployments are organized.

- **Model Training:**

1. Use AutoAI or a Python notebook to build a machine learning classification model.
2. The model learns patterns in applicant data to predict the correct NSAP scheme.

- **Evaluation:**

1. Review model accuracy, confusion matrix, and feature importance to ensure quality.
2. If needed, retrain or tune hyper parameters to improve performance.

- **Deployment:**

1. Promote the trained model to a Deployment Space and create a REST API.
2. This makes the model accessible for real-time predictions by external apps or users.

- **Prediction & Testing:**

1. Use the Test tab or API to input sample applicant data in JSON format.
2. The model instantly returns the predicted NSAP scheme name as output.

# SYSTEM APPROACH

The "System Approach" section outlines the overall strategy and methodology for developing and implementing the Predicting Eligibility for schemes. Here's a suggested structure for this section:

- **System requirements:**

- cloud platform:** IBM cloud lite

- Data Storage :** IBM cloud object storage

- Development Environment:** IBM watsonx.ai Studio

- AutoML:** IBM AutoAI

- **Library required to build the model:**

- Programming Language:** Python 3

- ML Libraries:** Pandas, scikit learn, matplotlib & seaborn

- Deployment:** IBM watson Machine Learning

- Input format:** JSON

- Output:** REST API Response

# Wow Factors & End users:

## Wow Factors:

### 1. AI-Driven Eligibility in Seconds

What typically takes hours of manual verification is reduced to **real-time scheme prediction** using AI.

### 2. Cloud-Based and API-Enabled

The entire solution is built and deployed using **IBM Cloud**, making it scalable, accessible from anywhere, and easily integrable into government systems.

### 4. Ready for Real-World Use

With its **REST API**, this project is not just theoretical but **deployable in live government offices, CSCs**

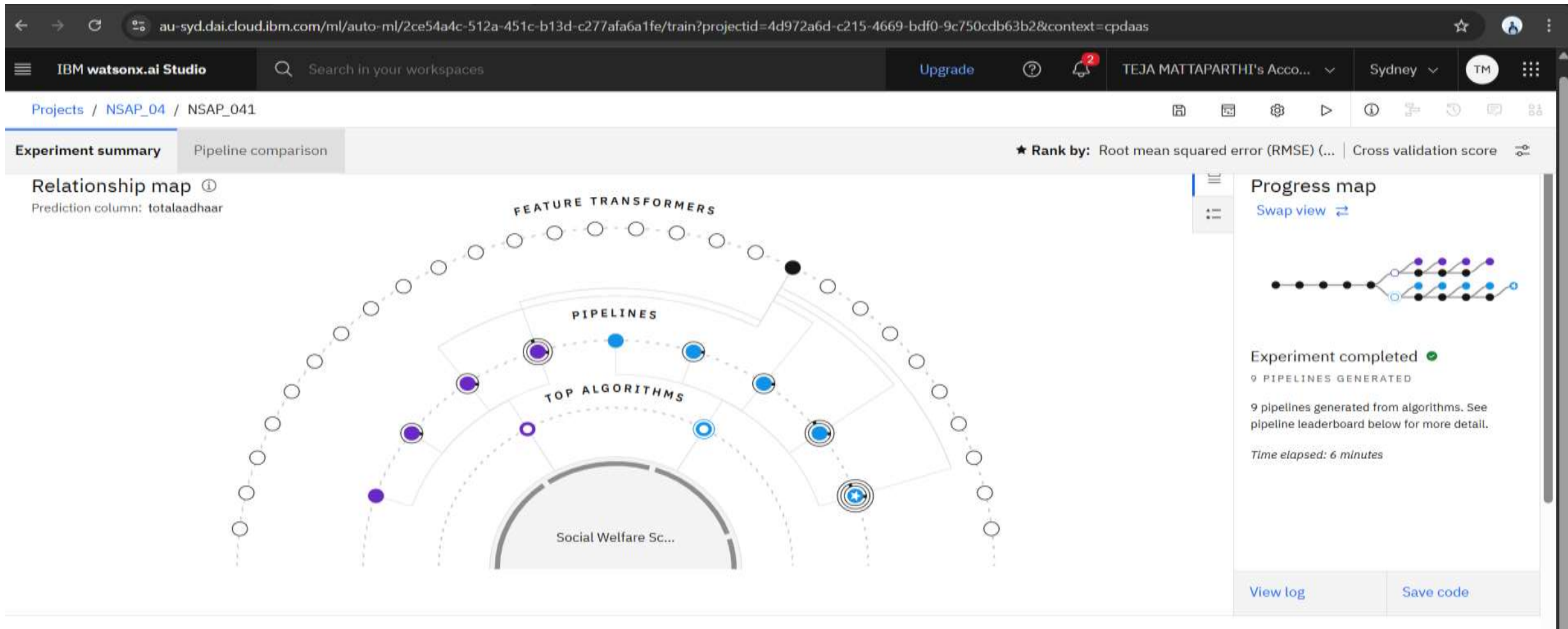
## End Users:

1. Government Officers
2. Village or ward Secretaries
3. NGOs & Social Workers
4. State/National Welfare Departments
5. Pension Applicants

# ALGORITHM & DEPLOYMENT






- In the Algorithm section, describe the machine learning algorithm chosen for predicting Eligibility schemes. Here's an example structure for this section:
- **Algorithm Selection:**
  - Random Forest
  - Logistic Regression
  - In AutoAI, Random Forest or Logistic Regressions typically performs best in terms of accuracy and F1-score.
- **Data Input:**
  - Specify The dataset includes socio-economic and demographic details of applicants across districts. It is in CSV format and contains both input features and the target label.
- **Training Process:**
  - IBM Watson Studio (AutoAI or Notebook).
- **Prediction Process:**
  - Detail The model is promoted to IBM Watson Machine Learning and deployed as a REST API.

# RESULTS



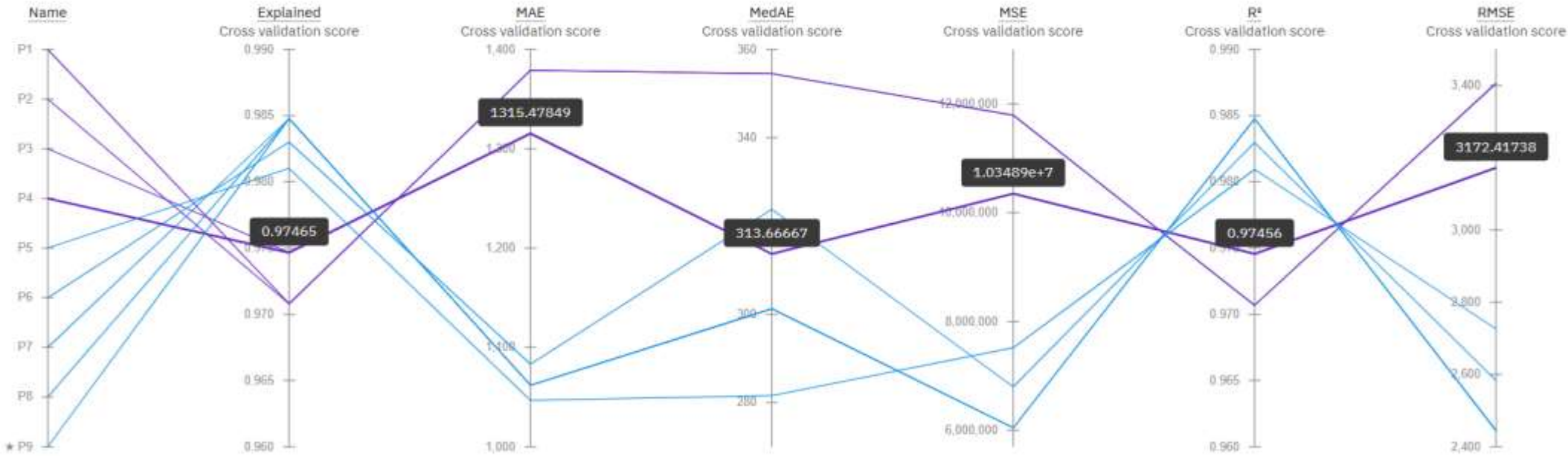


## Pipeline leaderboard

	Rank 	Name	Algorithm	Specialization	RMSE (Optimized) Cross Validation	Enhancements	Build time	
★	1	Pipeline 9	 Batched Tree Ensemble Regressor (Snap Boosting Machine Regressor)	INCR	2444.312	HPO-1 FE HPO-2 BATCH	00:01:27	
	2	<a href="#">Pipeline 8</a>	 Snap Boosting Machine Regressor		2444.312	HPO-1 FE HPO-2	00:01:24	<a href="#">Save as</a>
	3	Pipeline 7	 Snap Boosting Machine Regressor		2444.312	HPO-1 FE	00:01:17	
	4	Pipeline 6	 Snap Boosting Machine Regressor		2583.491	HPO-1	00:00:03	

Metric chart

Prediction column: totalaadhaar



Pipeline leaderboard

Rank	Name	Algorithm	Specialization	RMSE (Optimized) Cross Validation	Enhancements	Build time
------	------	-----------	----------------	--------------------------------------	--------------	------------

Projects / NSAP\_05 / Social Welfare Schemes.csv / Data Refinery



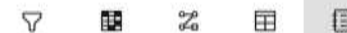
Steps

Use a code template to add a step



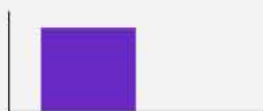
Data Profile Visualizations

Search Columns



finyear

Bar Pie



Mode 2025-2026  
2025-2026 2156

Missing Values

0%

0/2156

lgdstatecode

Histogram Box plot



Max 38  
Min 1

Missing Values

0%

0/2156

statename

Bar Pie



Mode UTTAR PRADESH  
UTTAR PRADESH 216

Missing Values

0%

0/2156

lgddistrictcode

Histogram Box plot



Max 763  
Min 1

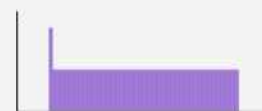
Missing Values

0%

0/2156

districtname

Bar Pie



Mode BIJAPUR  
BIJAPUR 6

Missing Values

0%

0/2156

schemecode

Bar Pie

totalbeneficiaries

Histogram Box plot

totalmale

Histogram Box plot

totalfemale

Histogram Box plot

totaltransgender

Histogram Box plot

## About this asset

### Name

Social Welfare Schemes.csv\_flow  
Data Refinery flow

### Description

What is the purpose of this Data Refinery flow?

### Asset details

Steps: 1

### Associated assets

- Source: Social Welfare Schemes.csv
- Target: Social Welfare Schemes\_c...

Last modified  
Not yet saved

Created on  
Not yet saved

[Projects](#) / [NSAP\\_05](#) / [Social Welfare Schemes.csv](#) / [Data Refinery](#)Steps 

Use a code template to add a step

Data Profile Visualizations

### SCATTER PLOT CHART

Off

Label field ⓘ

None

Primary title ⓘ

Subtitle ⓘ

Footnote ⓘ

XAxis Label ⓘ

YAxis Label ⓘ

**CHART TYPE**

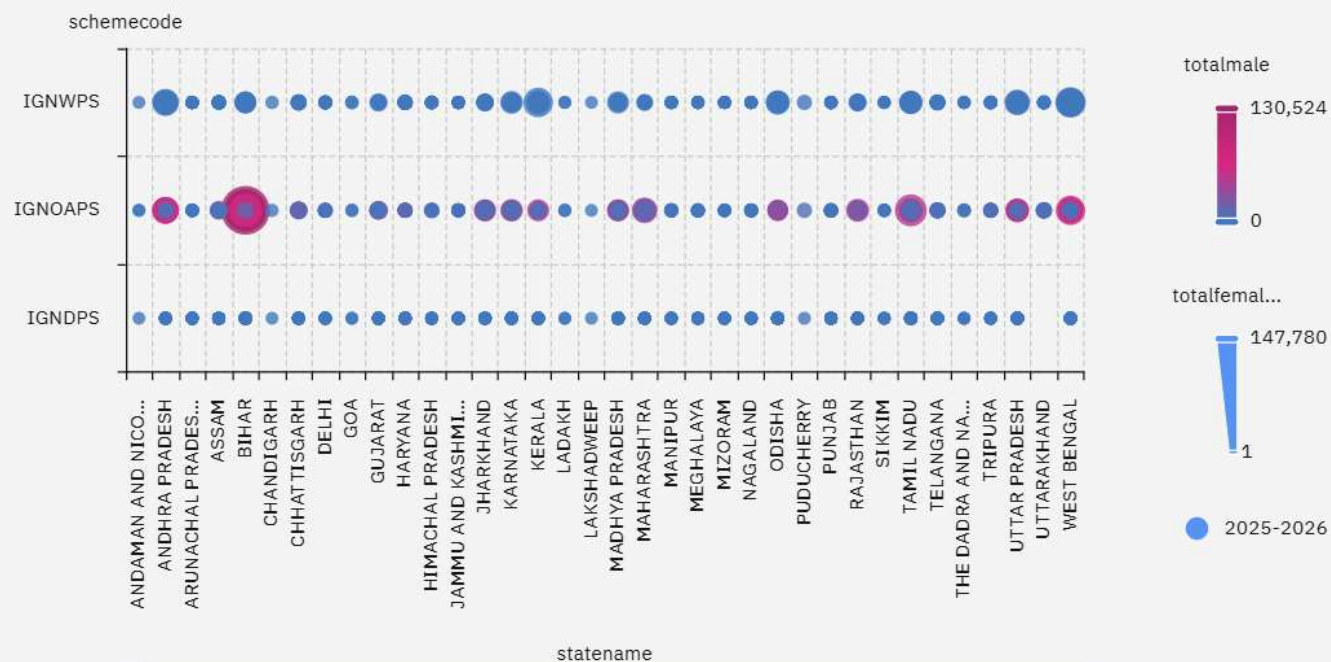
### Scatter plot

Line

### Multi-series

### Histogram

Population ...

ACTIONS 

### About this asset

Name

Social Welfare Schemes.csv\_flow  
Data Refinery flow

### Description

What is the purpose of this Data Refinery flow?

### Asset details

Steps: 1

### Associated assets

Source: Social Welfare Schemes.csv

Target: Social Welfare Schemes\_c...

Last modified

Not yet saved

Created on

Not yet saved

Projects / NSAP\_05 / Social Welfare Schemes.csv / Data Refinery

Steps

Use a code template to add a step

Data Profile Visualizations

PIE CHART

CHART TYPE

Scatter plot

Line

Multi-series

Histogram

Population ...

ACTIONS

Category

schemecode

3

Summary

- Count
- Count Unique
- Sum

Pie type

- Normal
- Ring
- Rose
- Rose Area
- Rose Ring
- Half Rose



IGNDPS IGNOAPS IGNWPS

About this asset

Name

Social Welfare Schemes.csv\_flow  
Data Refinery flow

Description

What is the purpose of this Data Refinery flow?

Asset details

Steps: 1

Associated assets

- Source: Social Welfare Schemes.csv
- Target: Social Welfare Schemes\_c...

Last modified

Not yet saved

Created on

Not yet saved



Steps Use a code template to add a step

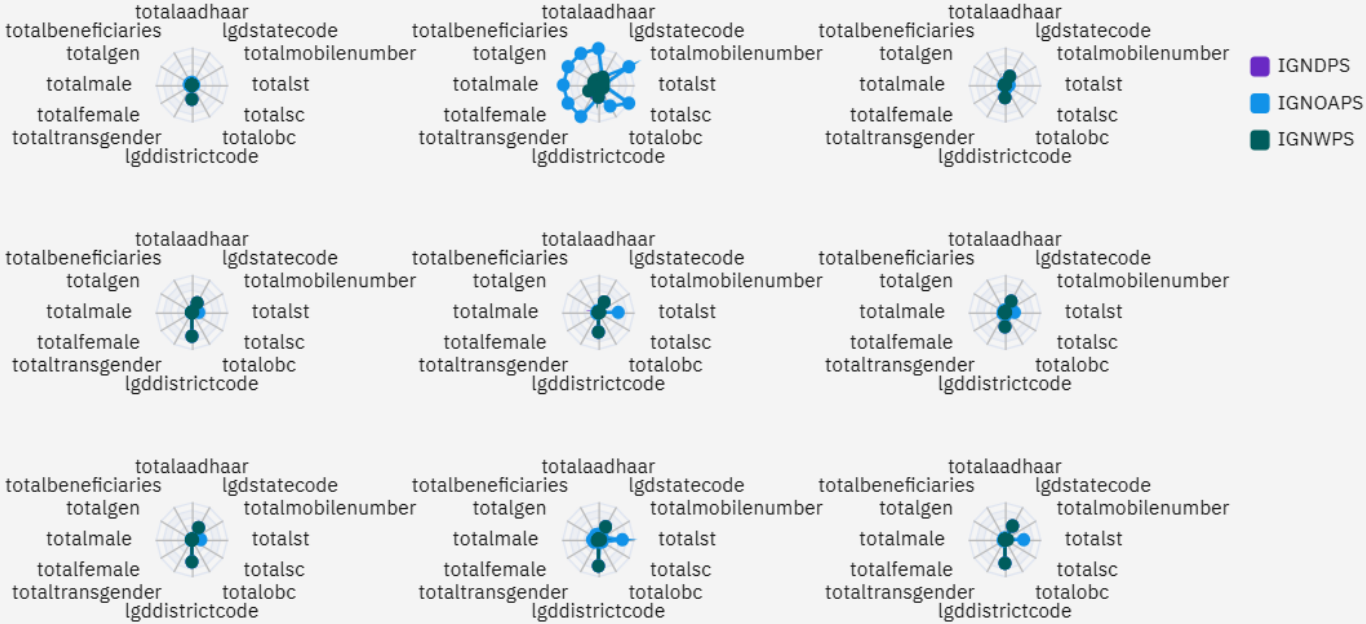
Data Profile Visualizations

RADAR CHART

- Columns\*
- totalaadhaar
  - totalbeneficiaries
  - totalgen
  - totalmale
  - totalfemale
  - totaltransgender
  - lgddistrictcode
  - totalobc
  - totalsc
  - totalst

CHART TYPE Scatter plot Line Multi-series Histogram Population ...

ACTIONS



About this asset

Name

Social Welfare Schemes.csv\_flow  
Data Refinery flow

Description

What is the purpose of this Data Refinery flow?

Asset details

Steps: 1

Associated assets

- Source: Social Welfare Schemes.csv
- Target: Social Welfare Schemes\_c...

Last modified

Not yet saved

Created on

Not yet saved

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

TEJA MATTAPARTHI's Acc...

Sydney

TM

Deployment spaces / NSAP\_06.2 / P9 - Snap Boosting Machine Regressor: NSAP\_06.1 /

NS

Prediction results

Close

X

Prediction type

Regression

Prediction distribution

Number of predictions

Prediction value

Display format for prediction results

☒ Table view ☐ JSON view

☐ Show input data ⓘ

	Prediction
1	3751.4080110710174
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	

Download JSON file

File Edit View Run Kernel Help

Trusted Memory:624 / 8192 MB

Code

Python 3.11

```
[2]: df.groupby("schemecode")["totalbeneficiaries"].sum()

df.groupby("districtname")["totalbeneficiaries"].sum().sort_values(ascending=False)
df.groupby("schemecode")[["totalmale", "totalfemale", "totaltransgender"]].sum()
```

[2]:

	totalmale	totalfemale	totaltransgender
schemecode			
IGNDPS	546833	296865	247
IGNOAPS	11350639	10689636	1751
IGNWPS	1	7005158	0

```
[3]: print(df.head())
print(df.info())
print(df.describe())
print(df.isnull().sum()) # Check for missing values
# Analyze the target variable distribution
print(df['schemecode'].value_counts())
# Replace 'NSAP_SCHEME_COLUMN' with actual column name
```

	fyear	lgdstatecode	statename	lgddistrictcode	districtname	\
0	2025-2026	1	JAMMU AND KASHMIR	1	ANANTNAG	
1	2025-2026	1	JAMMU AND KASHMIR	1	ANANTNAG	
2	2025-2026	1	JAMMU AND KASHMIR	1	ANANTNAG	
3	2025-2026	1	JAMMU AND KASHMIR	10	POONCH	
4	2025-2026	1	JAMMU AND KASHMIR	10	POONCH	

### Read data

Generate a code snippet to load data from a data asset or connection into your notebook.

Selected data

Social Welfare Schemes.csv

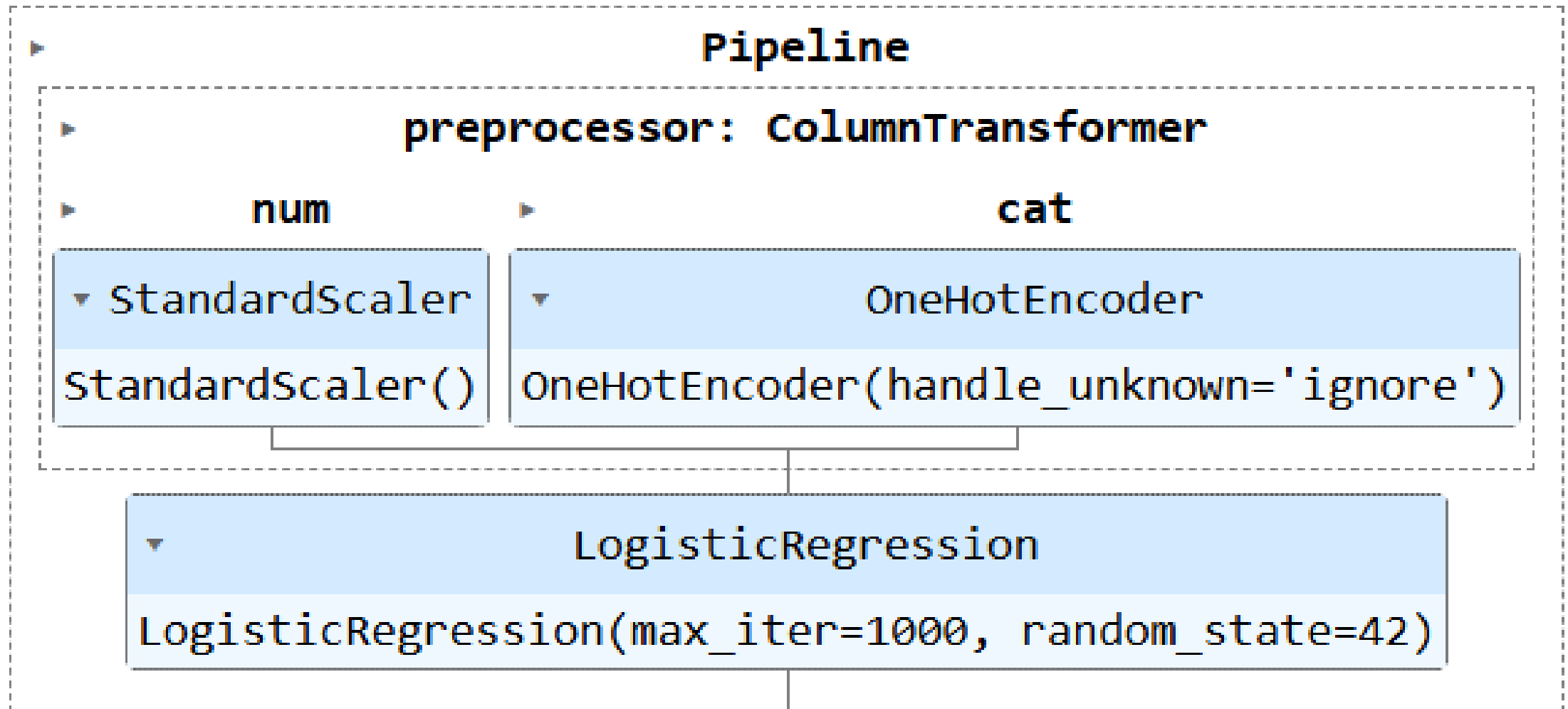
Load as

pandas DataFrame

Insert code to cell



[5]:

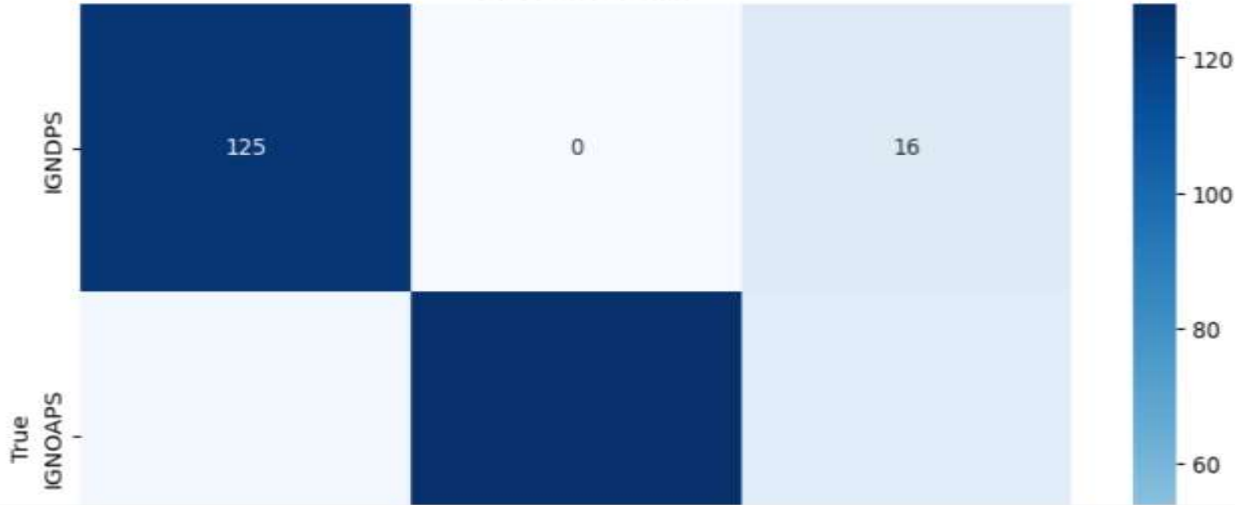


Accuracy: 0.8681

Classification Report:

	precision	recall	f1-score	support
IGNDPS	0.82	0.89	0.85	141
IGNOAPS	1.00	0.88	0.93	146
IGNWPS	0.80	0.84	0.82	145
accuracy			0.87	432
macro avg	0.88	0.87	0.87	432
weighted avg	0.88	0.87	0.87	432

Confusion Matrix



Read data

Generate a code snippet to load data from a data asset or connection into your notebook.

Selected data

Social Welfare Schemes.csv

Load as

pandas DataFrame

Insert code to cell

Projects / NSAP\_07 / NSAP\_07.1

↑ ▾ ⌂ ▾ ⬇ ⋮ &lt;/&gt; ⓘ ⌛ ⚙

File Edit View Run Kernel Help

🔍 + 🔗 📄 📄 ▶ ⏸ ↺ ⏭ Code ▾

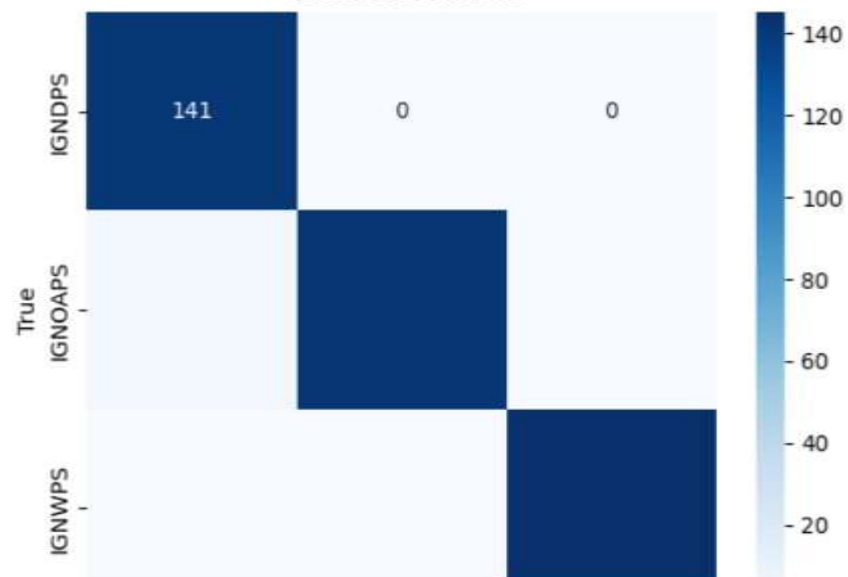
Trusted Memory:624 / 8192 MB ↗

Python 3.11 ○

Test Accuracy: 0.9907

	precision	recall	f1-score	support
IGNDPS	0.97	1.00	0.99	141
IGNOAPS	1.00	0.97	0.99	146
IGNWPS	1.00	1.00	1.00	145
accuracy			0.99	432
macro avg	0.99	0.99	0.99	432
weighted avg	0.99	0.99	0.99	432

Confusion Matrix



## Read data

Generate a code snippet to load data from a data asset or connection into your notebook.

Selected data

Social Welfare Schemes.csv



Load as:

pandas DataFrame ▾

Insert code to cell



# CONCLUSION

- Summarize This project successfully demonstrates how AI and IBM Cloud can automate the eligibility prediction process for NSAP schemes, replacing manual verification with fast, accurate, and reliable classification. By training a machine learning model on real-world socio-economic data, we ensure that **deserving individuals receive the right scheme on time**.

The deployed model, accessible via a REST API, offers a **practical, scalable, and ethical solution** that can be integrated into government platforms. It improves efficiency, transparency, and fairness in welfare delivery — marking a meaningful step toward **digitally empowered public service**.

# FUTURE SCOPE

- **Real time data streams:** Integrate the model from government databases or application portals
- **Explainable AI :** Incorporate tools like LIME or SHAP to provide a clear rationale for each prediction
- **Unstructured Data Analysis:** Expand the model's capabilities to include unstructured data from applicant forms
- **Continuous Learning:** Implement a system for model retraining and performance monitoring
- **Integration with Government Portals:** The deployed REST API endpoint is the key to integration
- **Mobile Application Development:** A mobile application could be developed for field workers
- **Fraud Detection:** The model could be re-purposed or enhanced to identify anomalies
- **Examples:** Mobile integration, multilingual input, periodic model retraining, expansion to other welfare schemes, fraud detection ad-on

# GITHUB LINK & REFERENCES

## GITHUB LINK:

[https://github.com/Teja1123-alt/Predict\\_Eligibility\\_For\\_NSAP\\_Schemes](https://github.com/Teja1123-alt/Predict_Eligibility_For_NSAP_Schemes)

## References:

- **IBM Cloud Lite**  
Cloud platform used for hosting, ML development, and deployment  
↗ <https://cloud.ibm.com/>
- **IBM Watson Studio**  
Machine Learning and Data Science platform used to build and train the model  
↗ <https://dataplatform.cloud.ibm.com>
- **Scikit-learn**  
Python machine learning library (used in manual modeling)  
↗ <https://scikit-learn.org/>
- **National Social Assistance Programme (NSAP)**  
Ministry of Rural Development, Government of India  
↗ <https://nsap.nic.in/>

## GETTING STARTED WITH ARTIFICIAL INTELLIGENCE:

In recognition of the commitment to achieve  
professional excellence



# TEJA MATTAPARTHI

Has successfully satisfied the requirements for:

## Getting Started with Artificial Intelligence



Issued on: Jul 15, 2025  
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/3aef7a06-8206-463c-a4b5-5b5ec5ab0274>



## JOURNEY TO CLOUD: ENVISIONING YOUR SOLUTION

In recognition of the commitment to achieve  
professional excellence



# TEJA MATTAPARTHI

Has successfully satisfied the requirements for:

## Journey to Cloud: Envisioning Your Solution



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

Verify: <https://www.credly.com/badges/5e66da36-49e5-4147-9f25-20eb2bb191c5>





## RETRIEVAL AUGMENTED GENERATION WITH LANGCHAIN

IBM **SkillsBuild**

Completion Certificate



This certificate is presented to  
**TEJA MATTAPARTHI**

for the completion of  
**Lab: Retrieval Augmented Generation with  
LangChain**

(ALM-COURSE\_3824998)

According to the Adobe Learning Manager system of record

**Completion date:** 16 Jul 2025 (GMT)

**Learning hours:** 20 mins



**THANK YOU**