
CAPSTONE PROJECT

PREDICTING ELIGIBILITY FOR : USING MACHINE LEARNING

Presented By:

Urbee Patra – Academy of Technology – Computer Science and Business Systems

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. This program consists of several sub-schemes for BPL citizens, 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.

This project 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

- The proposed system aims to address the challenge of predicting the right scheme for the BPL citizens. This involves leveraging data analytics and machine learning techniques to ensure that benefits are delivered to the right people efficiently.. The solution will consist of the following components:
- Data Collection:
 - Gather historical data on total male, total female, total transgender, and their socio-economic status.
 - Utilize data such as total aadhaar numbers and mobile phone numbers.
- 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 scheme selection.

- Machine Learning Algorithm:
 - Implement a machine learning algorithm, such as Decision Tree Classifier with enhancement of hyperparameter optimization and feature extraction to predict proper schemes based on historical patterns.
 - Consider incorporating other factors like which state and district to improve prediction accuracy.
- Deployment:
 - Develop a user-friendly interface or application that provides real-time predictions for bike counts at different hours.
 - Deploy the solution on a scalable and reliable platform, considering factors like server infrastructure, response time, and user accessibility.
- 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:

- Hardware-

Laptop (at least 8GB of RAM, an Intel i3 or AMD Ryzen 3 processor, and a 256GB SSD
Operating System: Windows 10 or 11, or macOS)

- Software-

IBM Cloud (watsonx.ai Studio service

20 capacity unit-hours (CUH) per month

50,000 tokens/data points per month

100 pages per month

Cloud object storage

Storage capacity up to 25 GB per month

Environment definition

large: 8 CPU and 32 GB RAM

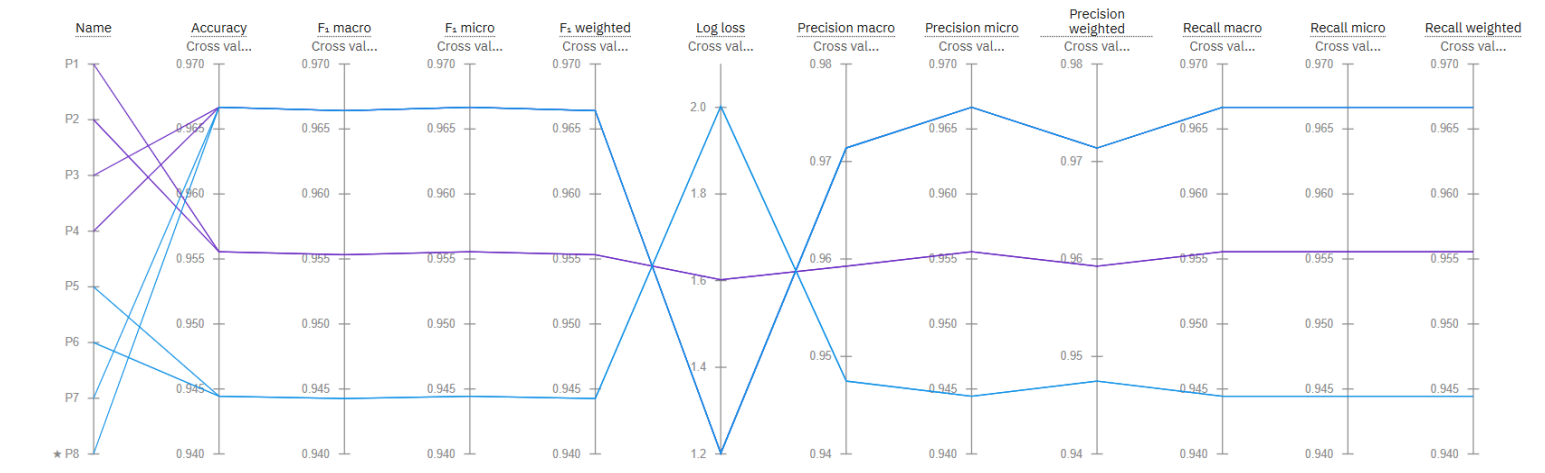
ALGORITHM & DEPLOYMENT

Algorithm Selection:

Experiment summary Pipeline comparison ★ Rank by: Accuracy (Optimized) | Cross valid:

Metric chart ⓘ

Prediction column: schemecode



Pipeline leaderboard ⓘ

	Rank	↑	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
★	1		Pipeline 8	Decision Tree Classifier	0.967	HPO-1 FE HPO-2	00:00:22
	2		Pipeline 7	Decision Tree Classifier	0.967	HPO-1 FE	00:00:19
	3		Pipeline 4	Snap Decision Tree Classifier	0.967	HPO-1 FE HPO-2	00:00:26
	4		Pipeline 3	Snap Decision Tree Classifier	0.967	HPO-1 FE	00:00:23

From the metric chart and pipeline leader board we can observe that Pipeline 8 with **Decision Tree Classifier** is best suited algorithm with **96.7% accuracy**. This was processed by IBM services with given input.

■ Data Input:

- The dataset was obtained from AI Kosh dataset, which is authenticated by Govt. of India.
- The name is **Districtwise Pension data under the National Social Assistance Programme NSAP**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	finyear	lgdstatco	statename	lgddistrict	districtnan	schemeco	totalbenef	totalmale	totalfemal	totaltrans	totalsc	totalst	totalgen	totalobc	totalaadha	totalmpbilenumber	
2	2025-2026	1	JAMMU AI	1	ANANTNA	IGNDPS	107	71	36	0	0	2	104	1	107	69	
3	2025-2026	1	JAMMU AI	1	ANANTNA	IGNOAPS	8393	5037	3356	0	37	232	8039	85	8327	7162	
4	2025-2026	1	JAMMU AI	1	ANANTNA	IGNWPS	203	0	203	0	1	15	181	6	201	160	
5	2025-2026	1	JAMMU AI	10	POONCH	IGNDPS	310	211	99	0	0	77	200	33	234	110	
6	2025-2026	1	JAMMU AI	10	POONCH	IGNOAPS	5959	3959	2000	0	2	1347	4368	242	3873	2287	

Add data source

Add files such as [tabular data \(CSV\)](#).

Browse
Select from project

DistrictwisePensiondataundertheNationalSocialAssistanceProgrammeNSAP.csv
Size: 9.57 KB Columns: 16

Configure details

Enable this option to predict future activity over a specified date/time range.
Yes No
Data must be structured and sequential. [Learn more](#)

What do you want to predict?
Prediction column ⓘ
schemecode

Prediction column: schemecode CUH remaining: 11.91 CUH

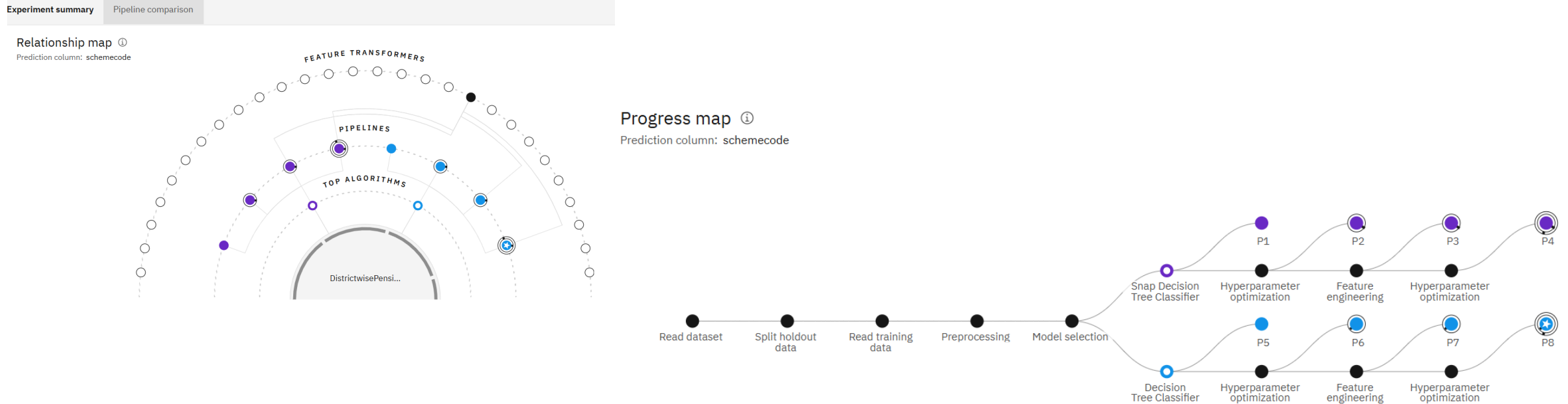
PREDICTION TYPE
Multiclass Classification
OPTIMIZED FOR
Accuracy & run time

Experiment settings
Run experiment

Data was feeded for
Multiclass
Classification and
find the required
Scheme Code

■ Training Process:

- Algorithm is trained using historical data. For training of data Preprocessing is done first. Then for each classifier hyperparameter tuning and feature extraction is done.
- As Decision Tree Classifier is Supervised Algorithm, so it is trained with labelled data.







■ Prediction process:

Enter input data

Text

JSON

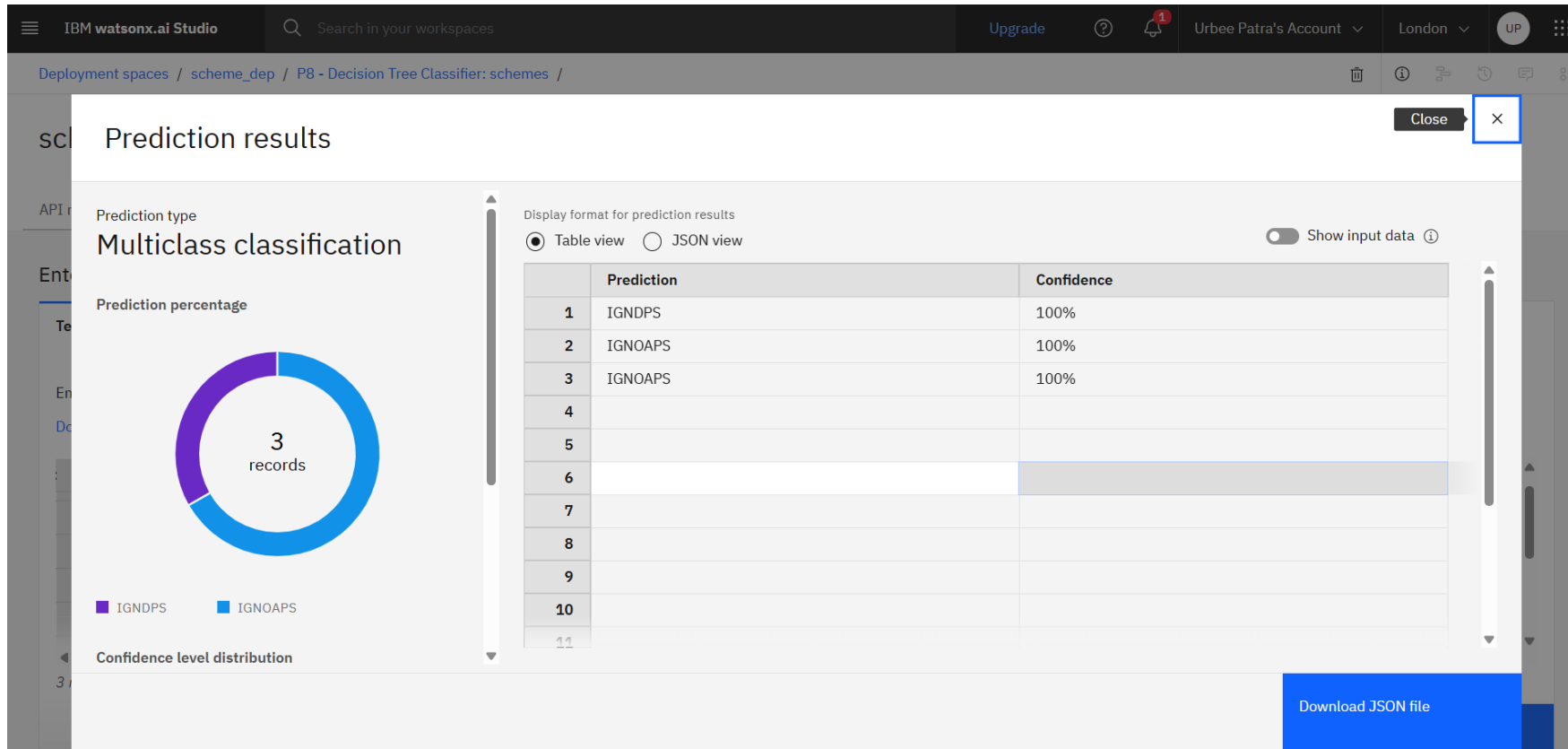
Enter data manually or use a CSV file to populate the spreadsheet. Max file size is 50 MB.

[Download CSV template](#)  [Browse local files](#)  [Search in space](#)  [Clear all](#) 

	finyear (other)	lgdstatecode (double)	statename (other)	lgddistrictcode (double)	districtname (other)	totalbeneficiaries (double)	totalmale (double)	totalfemale (double)
1	2025-2026	2	SIKKIM	12	GANGTOK	210	100	110
2	2025-2026	1	BIHAR	196	GAYA	1120	770	350
3	2025-2026	1	BIHAR	196	GAYA	350	125	213

These are test data for prediction. They are provided randomly by the developer. Then the ML model predicts the proper Scheme Code i.e the Scheme using Decision Tree Classifier.

RESULT



- This Multiclass Classification shows that for the given input data, the classifier provides appropriate predictions.
- It accurately predicts the most appropriate NSAP scheme for an applicant based on their demographic and socio-economic data.

CONCLUSION

- This model aims to provide financial assistance to the elderly, widows, and persons with disabilities belonging to below-poverty-line (BPL) households as the program consists of several sub-schemes, each with specific eligibility criteria.
- It is a reliable tool that could assist government agencies in quickly and accurately categorizing applicants, ensuring that benefits are delivered to the right people efficiently.

FUTURE SCOPE

- We can incorporate additional data sources from Govt. or from Private organisations,
- Optimize the algorithm for better performance. For example we can use SVM classifier.
- Try expanding the system to cover multiple cities or regions.
- Consider Integration of emerging technologies such as edge computing or advanced machine learning techniques.

REFERENCES

- S.R. Safavian, D. Landgrebe, “A survey of decision tree classifier methodology, 1991”, *IEEE Transactions on Systems, Man, and Cybernetics* (Volume: 21, Issue: 3, May-June 1991)
- B. Charbuty and A. Abdulazeez, “**Classification Based on Decision Tree Algorithm for Machine Learning**”, *JASTT*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
- AI Kosh dataset link –

https://aikosh.indiaai.gov.in/web/datasets/details/district_wise_pension_data_under_the_national_social_assistance_programme_nsap_1.html
- “IBM in the Cloud”, *International Journal of Micrographics & Optical Technology*, 2010, Vol 28, Issue 4/5, p4

IBM CERTIFICATIONS

CREDLY CERTIFICATE(GETTING STARTED WITH AI)

In recognition of the commitment to achieve
professional excellence



URBEE PATRA

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/b79e6235-40ac-4cd8-a94a-a4e904acdadb>



IBM CERTIFICATIONS

CREDLY CERTIFICATE(JOURNEY TO CLOUD)

In recognition of the commitment to achieve
professional excellence



URBEE PATRA

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution



Issued on: Jul 17, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/4b25f206-ce09-4c79-8501-be8250ba704b>



IBM CERTIFICATIONS

CREDLY CERTIFICATE(RAG LAB)

26/07/2025, 08:32

Completion Certificate | SkillsBuild

IBM **SkillsBuild**

Completion Certificate



This certificate is presented to

URBEE PATRA

for the completion of

**Lab: Retrieval Augmented Generation with
LangChain**

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 26 Jul 2025 (GMT)

Learning hours: 20 mins



THANK YOU