
CAPSTONE PROJECT

PREDICTING ELIGIBILITY FOR NSAP SCHEME USING MACHINE LEARNING

Presented By:

1. Aditya Sharma - MCKV Institute Of Engineering - CSE

OUTLINE

- Problem Statement
- 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

1 Overview

- Build a **multi-class classification model** using **IBM Cloud Lite AutoAI** to predict the correct NSAP scheme for each applicant.
- Automate the eligibility verification process to reduce delays and errors.

2 Steps in the Solution

1. **Data Collection** – Use AI Kosh *District-wise Pension Data* (demographic & socio-economic info).
2. **Data Preprocessing** – Clean data, handle missing values, encode categories, engineer features.
3. **Model Training with AutoAI** –
 - Select *Scheme Type* as the target variable.
 - AutoAI automatically tests multiple algorithms & preprocessing pipelines.
4. **Evaluation** – Use confusion matrix, per-class precision/recall to validate model.
5. **Deployment** – Publish as a REST API in IBM Cloud for real-time predictions.
6. **User Interface** – Simple input form for government staff to get instant scheme recommendations + explanation.

3 Key Advantages

- **Accuracy:** AI-driven eligibility prediction reduces human errors.
- **Speed:** Faster applicant processing → timely benefit delivery.
- **Transparency:** Model explanation shows why a scheme was suggested.
- **Scalability:** Can be used for all districts and updated with new data.

SYSTEM APPROACH

1 Concept

A systematic approach to automate NSAP scheme eligibility prediction using AI Kosh data and IBM AutoAI.

The system follows a **data → model → deployment → usage** pipeline.

2 Steps in the System

1. Input Layer (Data Acquisition)

- AI Kosh NSAP dataset with applicant demographic & socio-economic details.

- Includes attributes like

Finyear,Lgdstatecode,Statename,Lgddistrictcode,Districtname,Schemecode,Totalbeneficiaries,Totalmale,Totalfemale,Totaltransgender,Totalsc,Totalst,Totalgen,Totalobc,Totalaadh
aar,Totalmpbilenumber.

2. Processing Layer (Preprocessing & Feature Engineering)

- Data cleaning (remove duplicates, handle missing values).
- Encoding categorical variables.
- Feature creation .

3. Modeling Layer (Machine Learning)

- Use IBM AutoAI to train a **multi-class classification model**.
- Test multiple ML algorithms automatically (e.g. Decision tree classifier , Snap decision tree classifier).
- Select best-performing pipeline.

4. Deployment Layer (Cloud API)

- Deploy trained model in IBM Cloud.
- Make it available as a REST API for integration.

5. Application Layer (User Interface)

- Government staff inputs applicant details.
- Model predicts eligible NSAP scheme.
- Provide confidence score & explanation.

3 Advantages of System Approach

- Modular: Each stage can be improved independently.
- Scalable: Works for any number of applicants or districts.
- Transparent: Decision-making process can be reviewed.
- Fast & Accurate: Reduces manual effort and errors.

ALGORITHM & DEPLOYMENT

1. Algorithm Selection

- **Type:** Multi-Class Classification (via IBM AutoAI)
- **Best Model Chosen:** Decision Tree Classifier
- **Reason for Selection:**
 - Target is categorical (multiple NSAP schemes)
 - Tabular demographic & socio-economic data
 - AutoAI optimizes performance & explainability

2. Data Input

- Finyear
- Lgdstatecode
- Statename
- Lgddistrictcode
- Districtname
- Schemecode
- Totalbeneficiaries
- Totalmale
- Totalfemale
- Totaltransgender
- Totalsc
- Totalst
- Totalgen
- Totalobc
- Totalaadhaar
- Totalmpbilenumber

ALGORITHM & DEPLOYMENT

3. Training Process

- Data cleaning & encoding of categorical features
- Feature engineering
- **Stratified K-Fold Cross-Validation** to handle class imbalance
- Best pipeline selected & validated

4. Prediction Process

- Applicant details entered via UI or API
- Model outputs **Predicted Scheme(Schemecode)+ Confidence Score**
- Low-confidence cases flagged for manual review
- Real-time eligibility prediction for faster benefit delivery

RESULT

Model Performance

- Accuracy: ~96%

Visual Performance Metrics (Add Charts)

- Confusion Matrix – Shows actual vs predicted schemes.

Service Details - IBM Cloud IBM Watsonx.ai Studio NSAP eligibility prediction

eu-gb.dataplatform.cloud.ibm.com/ml/auto-ml/274a520-26ec-4ab4-a9f0-b9ed387697fd/train?projectid=a40feed5-9092-4a40-89e7-c00a8dcbb9d&context=cpdaas

IBM Watsonx.ai Studio Search in your workspaces Upgrade ? Aditya Sharma's Account London AS

Projects / project / NSAP

Pipeline details
Pipeline 8

Rank Accuracy (Optimized) Algorithm Enhancements Save as

Model viewer Model information Feature summary Evaluation Model evaluation Confusion matrix Precision recall Threshold

Multi-class

Observed	Predicted			
	IGNDPS	IGNOAPS	IGNWPS	Percent correct
IGNDPS	4	0	0	100.0%
IGNOAPS	0	3	0	100.0%
IGNWPS	0	0	3	100.0%
Percent correct	100.0%	100.0%	100.0%	100.0%

Less correct More correct

11:17 04-08-2025

Service Details - IBM Cloud nsap.deploy — nsap schema | +

eu-gb.dataplatform.cloud.ibm.com/ml-runtime/deployments/af054d43-9bcf-45de-9de6-57c73da5a816/test?space_id=21aaccac-d159-4474-a2a6-d1337f56be43&context=cpdaas&flush=true

IBM Watsonx.ai Studio Search in your workspaces Upgrade ? Aditya Sharma's Account London AS

Deployment spaces / nsap schema / P8 - Decision Tree Classifier: NSAP /

nsap Prediction results

Prediction type: Multiclass classification

Prediction percentage:

3 records

Prediction Confidence

1	IGNWPS	100%
2	IGNDPS	100%
3	IGNDPS	100%
4		
5		
6		
7		
8		
9		

Show input data Download JSON file

04-08-2025 10:37

Service Details - IBM Cloud IBM Watsonx.ai Studio

eu-gb.dataplatform.cloud.ibm.com/ml/auto-ml/274a520-26ec-4ab4-a9f0-b9ed387697fd/train?projectid=a40feed5-9092-4a40-89e7-c00a8dcbb9d&context=cpdaas

IBM Watsonx.ai Studio Search in your workspaces Upgrade ? Aditya Sharma's Account London AS

Projects / project / NSAP

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

Time elapsed: 2 minutes

DistrictwisePensi...

View log Save code

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	00:00:24
2	Pipeline 7	Decision Tree Classifier	0.967	HPO-1 FE	00:00:21
3	Pipeline 4	Snap Decision Tree Classifier	0.967	HPO-1 FE HPO-2	00:00:24
4	Pipeline 3	Snap Decision Tree Classifier	0.967	HPO-1 FE	00:00:21

ENG IN 04-08-2025 10:45

edunet foundation

SCREENSHOTS

The screenshot shows the IBM Cloud dashboard with a dark theme. At the top, there's a search bar and navigation links for Catalog, Manage, and account info. Below the header, a "For you" section displays six service cards:

- Build**: Explore IBM Cloud with this selection of easy starter tutorials and services.
- Track emissions with Carbon Calculator**: View estimated greenhouse gas emissions for your IBM Cloud account and export data for ESG reporting. (Recommended, 1 min)
- Use Watson Assistant**: Watson Assistant lets you build conversational interfaces into any application, device, or channel. (Popular, 2 min)
- Use Watson Studio**: Watson Studio provides a suite of tools and a collaborative environment for data scientists, developers and domain experts. (Popular, 2 min)
- Learn about API Keys**: Create API Keys to identify the application or user calling an API. Most IBM Cloud services require API Keys when calling their APIs. (Recommended, 5 min)
- Build with Watson**: Chatbots, insights, recognizers, and more. Explore the AI platform for business. (Popular, 3 min)

Below the cards, there are sections for "IBM Cloud status", "Recent support cases", "Planned maintenance", and "Total emissions". The "Total emissions" section includes a chart showing Emissions (kgCO₂e) over time. The bottom of the screen shows a taskbar with various icons and the date 04-08-2025.

The screenshot shows the "Service Details - IBM Cloud" page for "watsonx.ai Studio-ht". The page has a dark theme and includes the following sections:

- Resource list /**: Shows the resource name "watsonx.ai Studio-ht" and its tag "epdaas".
- Manage**: Includes a "Plan" section.
- watsonx.ai Studio in Cloud Pak for Data and watsonx**: A descriptive text block stating "Build and deploy machine learning models on either platform. Work with foundation models on watsonx as a Service." It features a "Launch in" button.
- Helpful links**: Links to Documentation, Learning path, and Videos.
- Diagram**: An architectural diagram showing "IBM watsonx.ai Studio in Cloud Pak for Data and watsonx" at the top, with "IBM Cloud Pak for Data, watsonx Unifying platforms" and "IBM Cloud Base cloud Infrastructure" below it.
- Bottom navigation**: Includes a taskbar with various icons and the date 04-08-2025.

The screenshot displays a workflow for generating machine learning pipelines in IBM Watson AI Studio.

Top Left Window: Shows the "Experiment summary" tab with a "Relationship map". The map illustrates the relationships between "FEATURE TRANSFORMERS", "PIPELINES", and "TOP ALGORITHMS". A progress bar indicates "8 PIPELINES GENERATED" and "Time elapsed: 2 minutes". Buttons for "View log" and "Save code" are present.

Top Right Window: Shows the "Pipeline details" for "Pipeline 8". The pipeline is ranked 1, has an accuracy of 1 (Holdout), and uses a "Decision Tree Classifier" algorithm with "HPO-1" and "FE" enhancements. A "Confusion matrix" table is displayed:

Observed	Predicted		
	IGNDPS	IGNOAPS	IGNWPS
IGNDPS	4	0	0
IGNOAPS	0	3	0
IGNWPS	0	0	3
Percent correct	100.0%	100.0%	100.0%

Bottom Window: Shows the "Pipeline leaderboard". It lists four pipelines based on accuracy (Optimized) and cross-validation score. Pipeline 8 is at the top with an accuracy of 0.967.

Rank	Name	Algorithm	Accuracy (Optimized) Cross Validation	Enhancements	Build time
1	Pipeline 8	Decision Tree Classifier	0.967	HPO-1 FE HPO-2	00:00:24
2	Pipeline 7	Decision Tree Classifier	0.967	HPO-1 FE	00:00:21
3	Pipeline 4	Snap Decision Tree Classifier	0.967	HPO-1 FE HPO-2	00:00:24
4	Pipeline 3	Snap Decision Tree Classifier	0.967	HPO-1 FE	00:00:21

The screenshot displays two browser windows side-by-side, illustrating the deployment process of a machine learning model.

Left Window: Input Schema

The URL is eu-gb.dataplatform.cloud.ibm.com/ml-runtime/models/60f497ff-53e6-48e3-8ce5-f6e8ac4fe2b7?project_id=a40feed5-9092-4a40-89e7-c00a8dc8b9d&context=cpdaas.

Input (1)

Column	Type
districtname	other
finyear	other
lgddistrictcode	double
lgystatecode	double
statename	other
totalaadhaar	double
totalbeneficiaries	double
totalfemale	double

About this asset

Name: P8 - Decision Tree Classifier: NSAP

Description: No description provided.

Asset Details:

- Type: wml-hybrid_0.1
- Model ID: 60f497ff-53e6-48e3-8ce5-f6e8ac4fe2b7
- Software specification: hybrid_0.1
- Hybrid pipeline software specifications: autoai-kb_r124.1-py3.11

Tags: Add tags to make assets easier to find.

Last modified: 23 hours ago by Aditya Sharma
Created on: Aug 3, 2025 by Aditya Sharma

Right Window: Deployment Spaces

The URL is eu-gb.dataplatform.cloud.ibm.com/ml-runtime/spaces/21aaccac-d159-4474-a2a6-d1337f56be43/deployments?context=cpdaas.

nsap schema

Deployments tab is selected.

Name	Type	Status	Asset	Asset type	Tags	Last modified
nsap_deploy	Online	Deployed	P8 - Decision Tree Classifier: NSAP	Model		22 hours ago Aditya Sharma (You)

Items per page: 20 | 1–1 of 1 items

The screenshot displays two windows of the IBM Watsonx.ai Studio interface. The top window shows the deployment details for 'nsap_deploy' under the 'nsap schema' workspace. It includes sections for 'Endpoints for scoring' (private and public), 'Code snippets' (cURL, Java, JavaScript, Python, Scala), and an 'About this deployment' panel with information like Name (nsap_deploy), Description (No description provided), Deployment ID (af054d43-9bcf-45de-9de6-57c73da5a816), and Software specification (hybrid_0.1). The bottom window shows the 'Test' interface for the 'nsap Deploy' endpoint, featuring an 'Enter input data' section with a table for inputting data. A large black arrow points from the 'nsap schema' workspace in the top-left towards the 'Test' interface in the bottom-right.

Service Details - IBM Cloud

eu-gb.dataplatform.cloud.ibm.com/ml-runtime/deployments/af054d43-9bcf-45de-9de6-57c73da5a816?space_id=21aacac-d159-4474-a2a6-d1337f56be43&context=cpdas

IBM watsonx.ai Studio

Deployment spaces / nsap schema / P8 - Decision Tree Classifier: NSAP /

nsap_deploy Deployed Online

API reference Test

Endpoints for scoring ⓘ

Private endpoint
https://private.eu-gb.ml.cloud.ibm.com/v4/deployments/af054d43-9bcf-45de-9de6-57c73da5a816/predict

Bearer <token> ⓘ IAM

Public endpoint
https://eu-gb.ml.cloud.ibm.com/ml/v4/deployments/af054d43-9bcf-45de-9de6-57c73da5a816/predictions?v=

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

Code snippets

cURL Java JavaScript Python Scala

NOTE: you must set \$API_KEY below using information retrieved from your IBM Cloud account (<https://eu-gb.dataplatform.cloud.ibm.com/docs/>)
export API_KEY=<your API key>

About this deployment

Name nsap_deploy

Description No description provided.

Deployment Details

Deployment ID: af054d43-9bcf-45de-9de6-57c73da5a816

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

P8 - Decision Tree Classifier: NSAP
faa53a74-3453-4859-9933-d4f6c48638c0

Service Details - IBM Cloud

eu-gb.dataplatform.cloud.ibm.com/ml-runtime/deployments/af054d43-9bcf-45de-9de6-57c73da5a816/test?space_id=21aacac-d159-4474-a2a6-d1337f56be43&context=cpdas&flush=true

IBM watsonx.ai Studio

Deployment spaces / nsap schema / P8 - Decision Tree Classifier: NSAP /

nsap_deploy 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 ✕

	finyear (other)	lgdstatecode (double)	statename (other)	lgddistrictcode (double)	districtname (other)	totalbeneficiaries (double)	totalmale (double)	totalfemale (double)
1	2025-2026	1	JAMMU AND KASH	13	SRINAGAR	607	0	607
2	2025-2026	1	JAMMU AND KASH	14	UDHAMPUR	26	19	7
3	2025-2026	5	WEST BENGAL	1	ASANSOL	5	5	0

3 rows, 15 columns

Predict

Service Details - IBM Cloud | nsap Deploy — nsap schema | +

eu-gb.dataplatform.cloud.ibm.com/ml-runtime/deployments/af054d43-9bcf-45de-9def-57c73da5a816/test?space_id=21aaccac-d159-4474-a2a6-d1337f56be43&context=cpdaas&flush=true

Gmail YouTube Maps Gemini for Application

IBM Watsonx.ai Studio Search in your workspaces Upgrade Aditya Sharma's Account London AS

Deployment spaces / nsap schema / P8 - Decision Tree Classifier: NSAP /

Prediction results

Prediction type Multiclass classification

Prediction percentage

3 records

IGNWPS IGNDPS

Display format for prediction results

Table view JSON view Show input data

	Prediction	Confidence
1	IGNWPS	100%
2	IGNDPS	100%
3	IGNDPS	100%
4		
5		
6		
7		
8		
9		

Download JSON file

Service Details - IBM Cloud | nsap Deploy — nsap schema | +

eu-gb.dataplatform.cloud.ibm.com/ml-runtime/deployments/af054d43-9bcf-45de-9def-57c73da5a816/test?space_id=21aaccac-d159-4474-a2a6-d1337f56be43&context=cpdaas&flush=true

Gmail YouTube Maps Gemini for Application

IBM Watsonx.ai Studio Search in your workspaces Upgrade Aditya Sharma's Account London AS

Deployment spaces / nsap schema / P8 - Decision Tree Classifier: NSAP /

Prediction results

API Ent Te En Do

Display format for prediction results

Table view JSON view Show input data

	Prediction	Confidence
1	IGNWPS	100%
2	IGNDPS	100%
3	IGNDPS	100%
4		
5		
6		
7		
8		
9		

Confidence level distribution

Number of records Confidence level

IGNWPS IGNDPS

Download JSON file

CONCLUSION

Findings

- The IBM AutoAI-based multi-class classification model achieved ~96% accuracy.
- Automated predictions significantly reduce manual verification time and human errors, enabling faster benefit delivery.

Effectiveness of the Proposed Solution

- **Reliable:** Consistently predicts correct NSAP scheme for majority of applicants.
- **Scalable:** Can be applied across districts with minimal modifications.
- **Transparent:** Provides feature importance and confidence scores for decisions.
- **Impact:** Helps ensure eligible beneficiaries receive timely financial support.

Challenges Encountered

- **Class Imbalance:** Some schemes had fewer samples, reducing recall.
- **Missing/Incomplete Data:** Required imputation and careful preprocessing.
- **Policy Rule Variations:** Differences in district-level eligibility rules needed standardization

FUTURE SCOPE

Potential Improvements

- Collect more balanced training data for underrepresented schemes.
- Integrate feedback loop from manual reviews to retrain model regularly.
- Add bias & fairness checks to ensure equitable decisions.
- Enhance UI integration for easier use by government staff.

REFERENCES

1. AI Kosh Dataset – District-wise Pension Data under NSAP

Government of India, National AI Portal

https://aikosh.indiaai.gov.in/web/datasets/details/district_wise_pension_data_under_the_national_social_assistance_programme_nsap_1.html

2. IBM AutoAI Documentation

IBM Cloud Docs – *Automating AI model selection, feature engineering, and hyperparameter tuning*

<https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/autoai-overview.html>

3. Brownlee, J. (2020). **Machine Learning Algorithms from Scratch**. Machine Learning Mastery.

Practical guide to understanding and implementing common ML classification algorithms.

4. Pedregosa, F. et al. (2011). **Scikit-learn: Machine Learning in Python**. *Journal of Machine Learning Research*, 12, 2825–2830.

<https://jmlr.org/papers/v12/pedregosa11a.html>

5. Chawla, N. V., et al. (2002). **SMOTE: Synthetic Minority Over-sampling Technique**. *Journal of Artificial Intelligence Research*, 16, 321–357.

Techniques for handling imbalanced datasets, relevant for underrepresented NSAP scheme classes.

6. Kumar, R., & Singh, A. (2019). **Applications of Machine Learning in Public Welfare Schemes**. *International Journal of Computer Applications*,

178(32), 25–30.

Explores ML use cases in government benefit distribution and eligibility prediction.

7. Han, J., Kamber, M., & Pei, J. (2011). **Data Mining: Concepts and Techniques**. Morgan Kaufmann.

Covers best practices in data preprocessing, feature engineering, and evaluation.

IBM CERTIFICATIONS

The screenshot shows a Microsoft Edge browser window with a dark theme. The address bar displays the file path: C:/Users/adity/Downloads/IBMDesign20250724-34-rj53s4.pdf. The main content area shows a digital certificate for "Getting Started with Artificial Intelligence". The certificate includes the following text:
In recognition of the commitment to achieve professional excellence
Aditya Sharma
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/0fa9306e-0b54-4c2e-ab8d-8fc119a1f9fc>
The certificate is framed by a blue border and has a white background. The IBM logo is visible in the bottom right corner of the certificate frame.

IBM CERTIFICATIONS

The screenshot shows a Microsoft Edge browser window with a dark theme. The address bar displays 'IBMDesign20250724-29-w8wf2.pdf'. The main content area shows a digital badge for 'Journey to Cloud: Envisioning Your Solution' issued to Aditya Sharma. The badge is white with a blue border and features the IBM logo at the top right. The text on the badge reads: 'In recognition of the commitment to achieve professional excellence', 'Aditya Sharma', 'Has successfully satisfied the requirements for: Journey to Cloud: Envisioning Your Solution', 'Issued on: Jul 19, 2025', 'Issued by: IBM SkillsBuild', and includes a QR code and a verify link: <https://www.credly.com/badges/3e5e72a4-50f9-4b94-8632-02ddf0d85dbd>. The bottom of the screen shows the Windows taskbar with various pinned icons.

IBM CERTIFICATIONS

The screenshot shows a completion certificate for Aditya Sharma. The certificate is titled "Completion Certificate" and is issued by "IBM SkillsBuild". It features a yellow ribbon icon. The text states that the certificate is presented to Aditya Sharma for the completion of the course "Lab: Retrieval Augmented Generation with LangChain" (ALM-COURSE_3824998). According to the Adobe Learning Manager system of record, the completion date was 24 Jul 2025 (GMT) and the learning hours were 20 mins. The certificate is displayed in a Microsoft Edge browser window on a Windows 10 desktop.

Completion Certificate

IBM SkillsBuild

This certificate is presented to

Aditya Sharma

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

Completion Certificate_ SkillsBuild.pdf

C:/Users/adity/Downloads/Completion%20Certificate%20_%20SkillsBuild.pdf

Import favorites | Gmail | YouTube | Maps | Amazon.co.uk - Only... | Express VPN | McAfee Security | LastPass password...

Draw | Ask Copilot

1 of 1

Edit with Acrobat

ENG IN 11:27 04-08-2025



THANK YOU