CAPSTONE PROJECT

NSAP Female Beneficiary Predictor using IBM AutoAI

Presented By:
Vanshika Wadhwa- Dyal Singh College, University of Delhi - Computer Science

OUTLINE

- **Problem Statement 34: NSAP Eligibility Predictor**
- **Proposed System/Solution**
- **System Development Approach**
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

edunet
foundation

Problem Statement

The National Social Assistance Programme (NSAP) provides pension support to eligible citizens. However, estimating scheme coverage (especially female beneficiaries) for planning and analysis is challenging.

Goal: Use ML to predict the number of female beneficiaries based on district and scheme-level indicators.
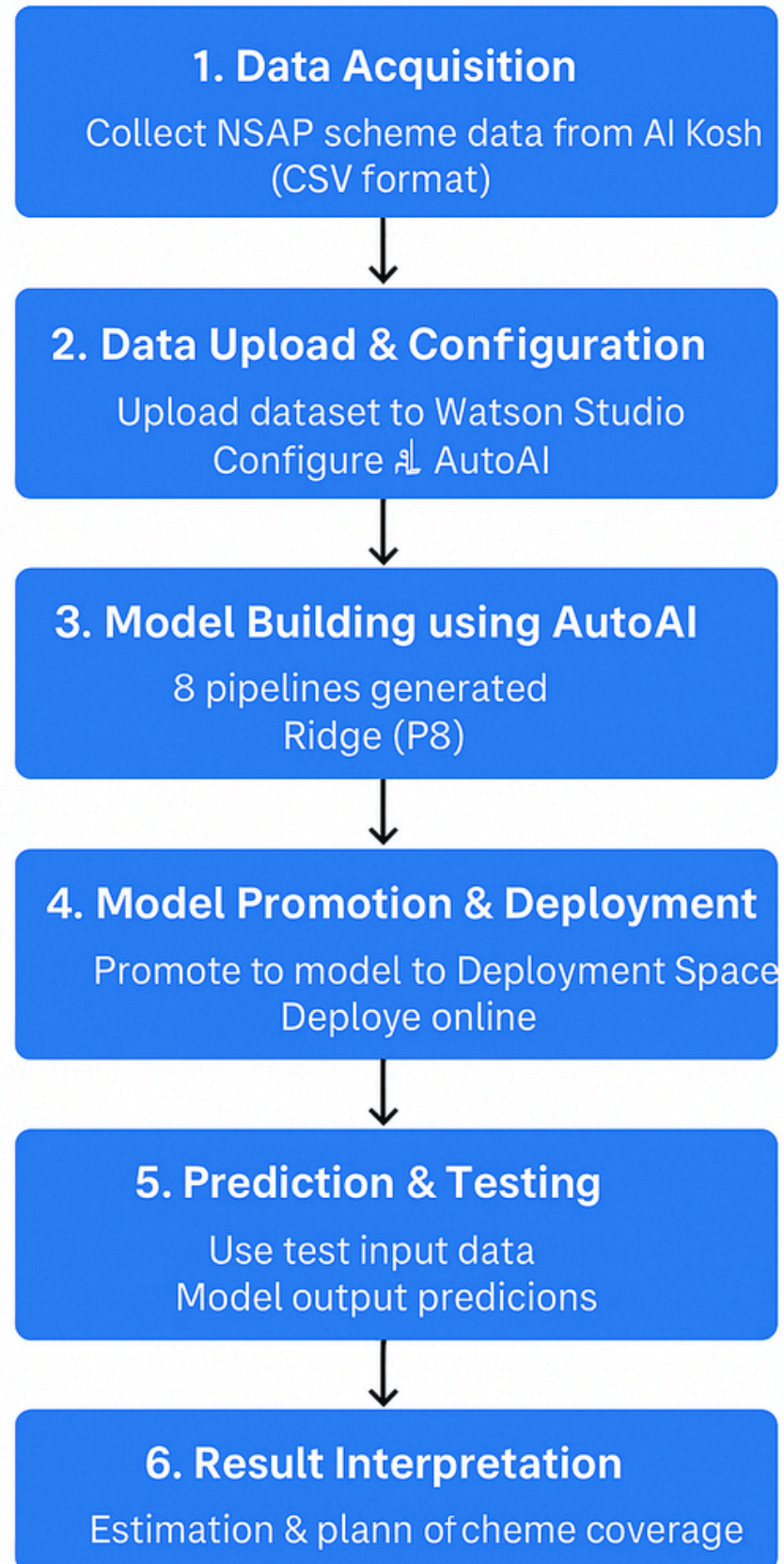
Proposed Solution

To address the challenge of estimating female beneficiaries under NSAP schemes across different Indian districts and states, we propose a Machine Learning-based regression solution built entirely using IBM Watson AutoAI.

◆ First, we sourced real scheme-level data from AI Kosh, containing demographic and administrative features such as:
- State/District codes
- Aadhaar counts
- Caste-wise population (SC/ST/OBC)
- Gender-wise totals
- Scheme codes, etc.

◆ This data was uploaded into IBM Watson Studio to create an AutoAI experiment.

◆ Using IBM AutoAI, a regression model was automatically trained to predict the `totalfemale` (number of female beneficiaries) based on the other columns.

◆ AutoAI selected the best pipeline (Ridge Regressor) after evaluating 8 different algorithmic pipelines with automatic:
- Preprocessing
- Feature Engineering
- Hyperparameter Optimization

◆ The model was then deployed to IBM Cloud using the in-built Deployment Space.

◆ Finally, real-time test data was provided through the UI, and the deployed model returned predictions with high accuracy.

This solution makes scheme planning and beneficiary forecasting automated, accurate, and cloud-based — perfect for policy planners or digital governance tools.

edunet
foundation

System Approach

1. Data Acquisition

Collect NSAP scheme data from AI Kosh (CSV format)

2. Data Upload & Configuration

Upload dataset to Watson Studio
Configure AutoAI

3. Model Building using AutoAI

8 pipelines generated
Ridge (P8)

4. Model Promotion & Deployment

Promote to model to Deployment Space
Deploye online

5. Prediction & Testing

Use test input data
Model output predicions

6. Result Interpretation

Estimation & plann of cheme coverage

Algorithm & Deployment

Algorithm Selection:

- IBM AutoAI evaluated 8 ML models and selected Ridge Regression based on lowest RMSE. Ridge handles multicollinearity well, making it suitable for structured NSAP scheme data.

Data Input:

- Key features used:
- Demographics: totalgen, totalobc, totalst, totaltransgender
- IDs: lgdstatecode, lgddistrictcode, schemecode
- Others: totalaadhaar, totalmobilenumber, totalbeneficiaries
- Target variable: totalfemale

Training Process:

- AutoAI automated:
- Data preprocessing
- Feature engineering
- Hyperparameter tuning
- Final model: Pipeline P8 – Ridge Regressor

Prediction Process:

- Model deployed to IBM Cloud.
- New data is entered via CSV/JSON → Model predicts totalfemale beneficiaries instantly via API.
- Let me know if you want a slide version with bullet points or icons next.

edunet
foundation

Result

Model Performance:

The final model selected by AutoAI was Ridge Regression (Pipeline P8).
 It achieved high accuracy in predicting totalfemale beneficiaries based on district-level input features.

- Evaluation Metric: Root Mean Squared Error (RMSE)
- Model RMSE: 9.18 (on holdout set)
- R^2 Score (Goodness of Fit): 0.92

This indicates that 92% of the variance in female counts can be explained by the model — showing strong prediction capability.

IBM watsonx.ai Studio

Search in your workspaces

Upgrade

Vanshika .'s Account

London

Projects / NSAP / Nsap

Configure AutoAI experiment

Nsap ✎

Autosaved: 5:45:19 PM

Browse | Select from project

nsapallschemes.csv
Size: 181.56 KB   Columns: 16

What do you want to predict?

Prediction column ⓘ

totalfemale ✕ ⌄

Prediction column: totalfemale

CUH remaining: 17.69 CUH

PREDICTION TYPE
Regression

OPTIMIZED FOR
RMSE & run time

Experiment settings ⚙   Run experiment ▷

---

eu-gb.dataplatform.cloud.ibm.com/ml/auto-ml/58a76971-5606-4bce-9453-96404199e2ae/train?projectid=f992db7a-1166-45b7-9f8b-903a5bddc5c5&cont...

IBM watsonx.ai Studio

Search in your workspaces   Upgrade   Vanshika .'s Account   London

Projects / NSAP / Nsap

Experiment summary | Pipeline comparison

★ Rank by: Root mean squared error (RMSE) (... | Cross validation score

Progress map ⓘ
Prediction column: totalfemale

Read dataset — Split holdout data — Read training data — Preprocessing — Model selection

Snap Decision Tree Regressor — Hyperparameter optimization — Feature engineering — Hyperparameter optimization
P1 P2 P3 P4

Ridge — Hyperparameter optimization — Feature engineering — Hyperparameter optimization
P5 P6 P7 P8

Relationship map
Swap view ⇄

Experiment completed ✔
8 PIPELINES GENERATED

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

Time elapsed: 3 minutes

View log   Save code

Pipeline leaderboard ▽

IBM watsonx.ai Studio

Search in your workspaces   Upgrade   Vanshika .'s Account   London

/ NSAP / Nsap

Configure AutoAI experiment

Nsap ✎

Autosaved: 5:45:19

Browse | Select from project

nsapallschemes.csv
Size: 181.56 KB   Columns: 16

What do you want to predict?

Prediction column ⓘ

totalfemale ✕ ⌄

Prediction column: totalfemale

CUH remaining: 17.69 CUH

PREDICTION TYPE
Regression

OPTIMIZED FOR
RMSE & run time

Experiment settings ⚙   Run experiment ▷

---

eu-gb.dataplatform.cloud.ibm.com/ml-runtime/deployments/a0fbf97d-a7b6-46c2-8774-1cbc44e2458b/test?space_id=ba0be3f4-2ab0-4bfa-8405-5a75dcb1...

IBM watsonx.ai Studio   Search in your workspaces   Upgrade   Vanshika .'s Account   London

Deployment spaces / NSAP / P8 - Ridge: Nsap

Prediction results

Prediction type
Regression

Prediction distribution

Display format for prediction results
◉ Table view   ◯ JSON view

Show input data ⓘ

	Prediction
1	98.880615234375
2	
3	
4	
5	
6	
7	

Prediction distribution:
Number of predictions / Prediction value
98.88061   98.88061

Download JSON file

Conclusion

- This project successfully demonstrated the use of IBM AutoAI to build a regression model that predicts the number of female beneficiaries under NSAP schemes using district-level demographic and administrative data.
- The model achieved high prediction accuracy with an R^2 score of 0.92, showing strong alignment between actual and predicted values. It was deployed on IBM Cloud, enabling real-time predictions through a secure and scalable API.

Effectiveness of the Solution:
- Automates scheme data analysis, reducing manual work
- Supports data-driven planning at district/state level
- Can improve transparency and coverage in welfare delivery

Challenges Faced:
- Limited data granularity (e.g., no household-level features)
- Minor inconsistencies in input fields across states
- Feature selection was entirely automated, limiting custom tuning

Conclusion

Potential Improvements:

- Integrate more features like literacy rate, employment, rural population
- Build a front-end UI (Streamlit or React) for field officers
- Expand model to classify eligibility for specific NSAP schemes

Why It Matters:

Accurate prediction of welfare scheme reach (e.g., female beneficiaries) is essential for:

- Ensuring equitable distribution of funds
- Reducing leakages or overlaps in scheme coverage
- Empowering policy-makers with data-backed decisions

This ML solution lays the foundation for smarter governance tools in public welfare systems.

edunet
foundation

Future scope

- **Add More Data:** Integrate census info, literacy rate, rural/urban population for richer predictions
- **Improve Model:** Use ensemble methods or XAI for better accuracy and interpretability
- **Scale to More Regions:** Expand coverage across all states and other welfare schemes
- **Tech Integration:** Deploy on edge devices, and build a chatbot for real-time field use

Goal: Transform into a smart welfare analytics tool for data-driven governance.

References

- AI Kosh – NSAP Dataset Government of India Open Dataset Portal
  https://aikosh.indiaai.gov.in
- NSAP Scheme Overview – Ministry of Rural Development, Govt. of India
  https://nsap.nic.in
- IBM Watson AutoAI Documentation –
  AutoAI: Automating machine learning model creation and deployment
  https://www.ibm.com/cloud/watson-studio/autoai

edu**net**
foundation

IBM Certifications

In recognition of the commitment to achieve professional excellence

Vanshika Wadhwa

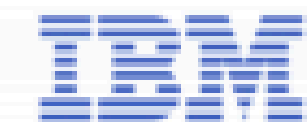Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 20, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/17d9e383-9b24-459a-8c85-318786572305

IBM

IBM Certifications

In recognition of the commitment to achieve professional excellence

Vanshika Wadhwa

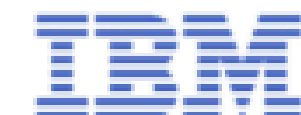Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 20, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/a55de007-def2-4551-b9a6-73ee7c7968bf
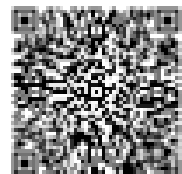
IBM **SkillsBuild**   Completion Certificate

This certificate is presented to

Vanshika Wadhwa

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

Github Repo: *https://github.com/vanshika-wadhwa/nsap_female_beneficiary_predictor*

THANK YOU

edunet
foundation