CAPSTONE PROJECT

PREDICTING ELIGIBILITY FOR NSAP SCHEMES USING MACHINE LEARNING

Presented By:

NAME: MUKILAN K
COLLEGE NAME: K RAMAKRISHNAN COLLEGE OF TECHNOLOGY
DEPARTMENT: ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

edunet foundation

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 NSAP provides financial support to vulnerable groups like the elderly, widows, and persons with disabilities from BPL households.
- Manually verifying applications and assigning schemes is slow, inconsistent, and prone to errors.
- A key challenge is accurately identifying the right scheme for each applicant based on their socio-economic and demographic data.

PROPOSED SOLUTION

To streamline the allocation process under the NSAP, a predictive machine learning model is developed using IBM Cloud services. The solution includes the following key components:

1. Data Collection

- Dataset sourced from the AI Kosh portal containing district-wise pension data under NSAP.
- Features include demographic and socio-economic details relevant to scheme eligibility (26 features used).

2. Data Preprocessing

- Cleaned missing or inconsistent entries to ensure data quality.
- Applied feature engineering to enhance prediction efficiency and model accuracy.
- Encoded categorical variables and normalized inputs for optimal training.

3. Machine Learning Model

- Utilized Random Forest Classifier to perform multi-class classification.
- Target column: schemecode indicating the appropriate NSAP scheme for an applicant.
- Model trained on 216 evaluation instances.

PROPOSED SOLUTION

4. Model Deployment

- Deployed on IBM Cloud using AutoAI and Snap ML tools.
- Allows real-time prediction of eligible scheme based on applicant input.
- Easily accessible by government officers via a streamlined cloud dashboard.

5. Model Evaluation

- Achieved high-performance metrics:
- Accuracy: 97.7% (Holdout), 98.4% (Cross-validation)
- Precision / Recall / F1-score (macro & weighted): All above 97%
- Log Loss: 0.222 (Holdout), 0.143 (Cross-validation)

SYSTEM APPROACH

1. System Requirements

- **IBM Cloud Lite Account** with Watson Studio and Object Storage
- **IAM API Key** for Watson Machine Learning deployment
- **Watson Studio Project** configured with a Python 3.x runtime
- **Access to AutoAI** for automated model training and pipeline generation
- **REST API Endpoint** for real-time scoring of deployed model

2. Libraries & Tools

- **pandas**, **numpy**– Data loading, cleaning, feature engineering
- **scikit-learn**– Preprocessing utilities (LabelEncoder, StandardScaler), evaluation metrics
- **ibm-watson-machine-learning** - Authentication, deployment, scoring via IBM Cloud APIs
- **ipywidgets**, **IPython.display**– Simple in-notebook UI for user inputs and result display
- **requests**– HTTP calls to the Watson ML scoring endpoint

SYSTEM APPROACH

3. Workflow Steps

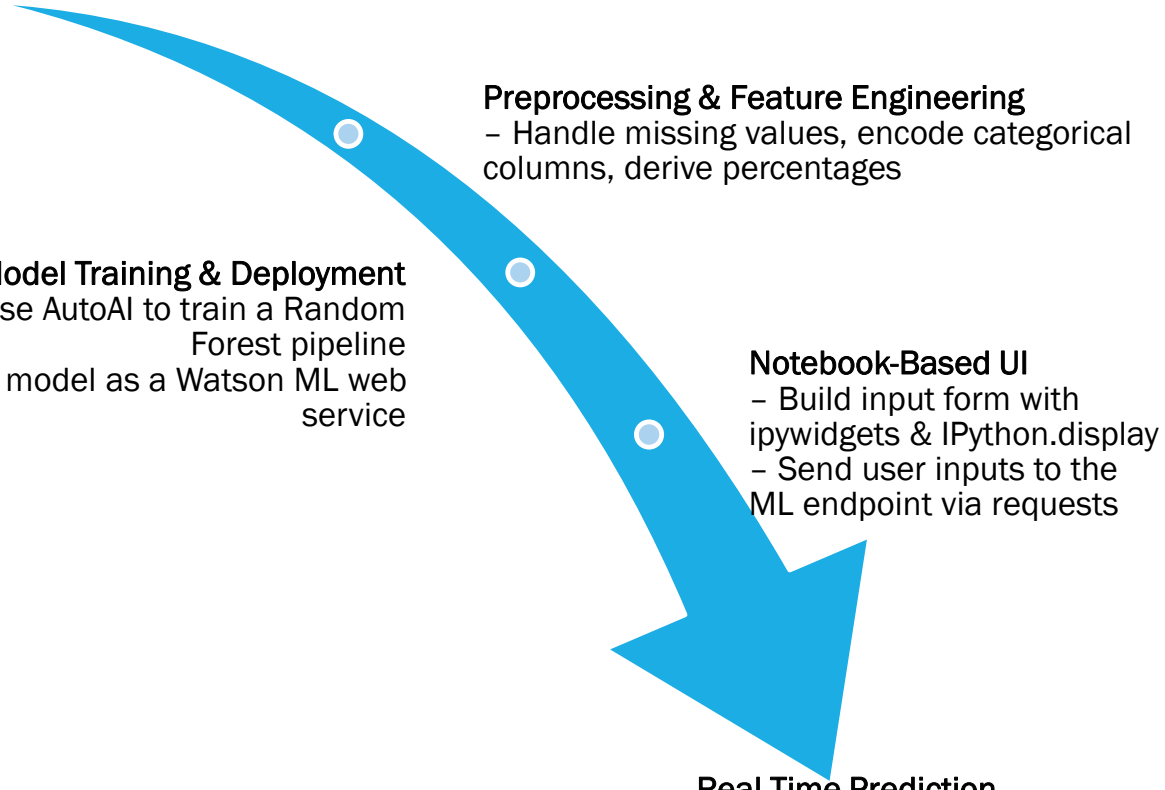
Data Ingestion
– Load CSV from object storage into pandas DataFrame

Preprocessing & Feature Engineering
– Handle missing values, encode categorical columns, derive percentages

Model Training & Deployment
– Use AutoAI to train a Random Forest pipeline
– Deploy model as a Watson ML web service

Notebook-Based UI
– Build input form with ipywidgets & IPython.display
– Send user inputs to the ML endpoint via requests

Real-Time Prediction
– Parse and present the predicted schemecode instantl



edunet foundation

ALGORITHM & DEPLOYMENT

Algorithm

1.Selection

- **Random Forest Classifier**: Robust for multi-class tabular data; handles mixed feature types and reduces overfitting.

2.Inputs

- 26 features including demographics (gender, SC/ST/OBC/general counts), Aadhaar/mobile coverage, geographic codes, and total beneficiaries.

3.Training

- **80/20 split** with 5-fold cross-validation
- Hyperparameter tuning (trees, depth) via AutoAI

4.Prediction

- New applicant data formatted into the same feature vector
- Model returns the highest-probability schemecode

Deployment

1.AutoAI → Watson ML

- Deployed directly from AutoAI as a REST-accessible service

2.Authentication

- IAM API key → Bearer token via IBM Cloud Identity

3.Scoring

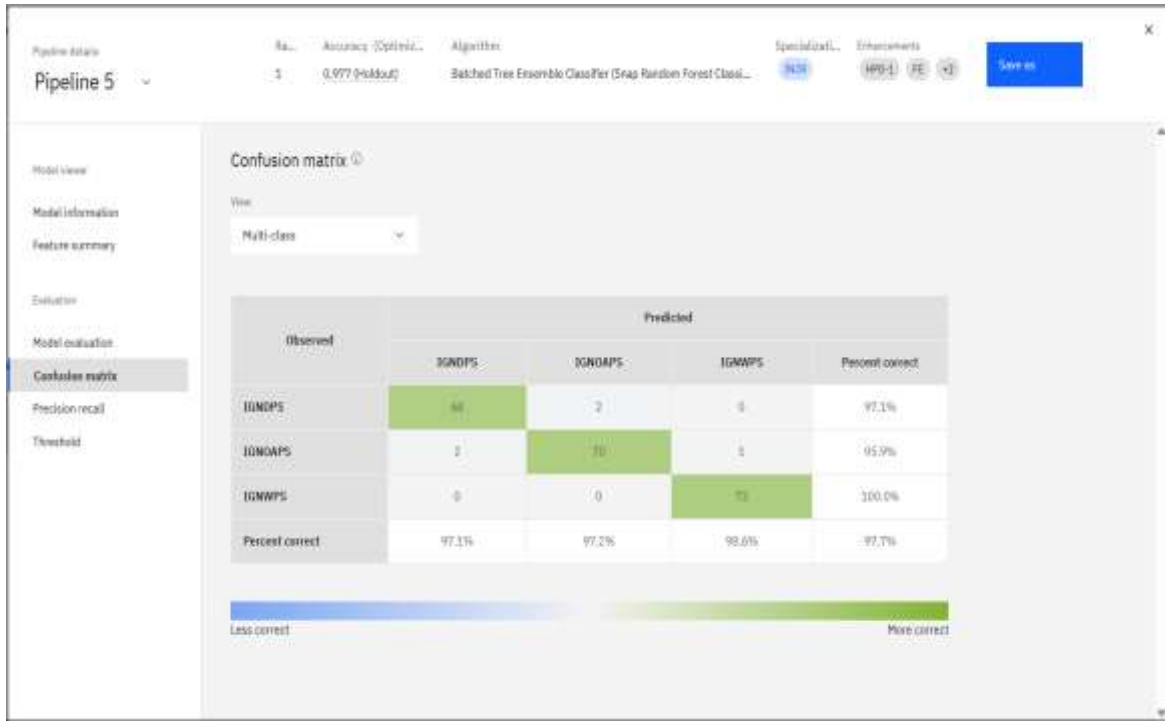
- Send JSON (fields + values) to endpoint using requests
- Receive predicted scheme in the response

4.Scalability & Monitoring

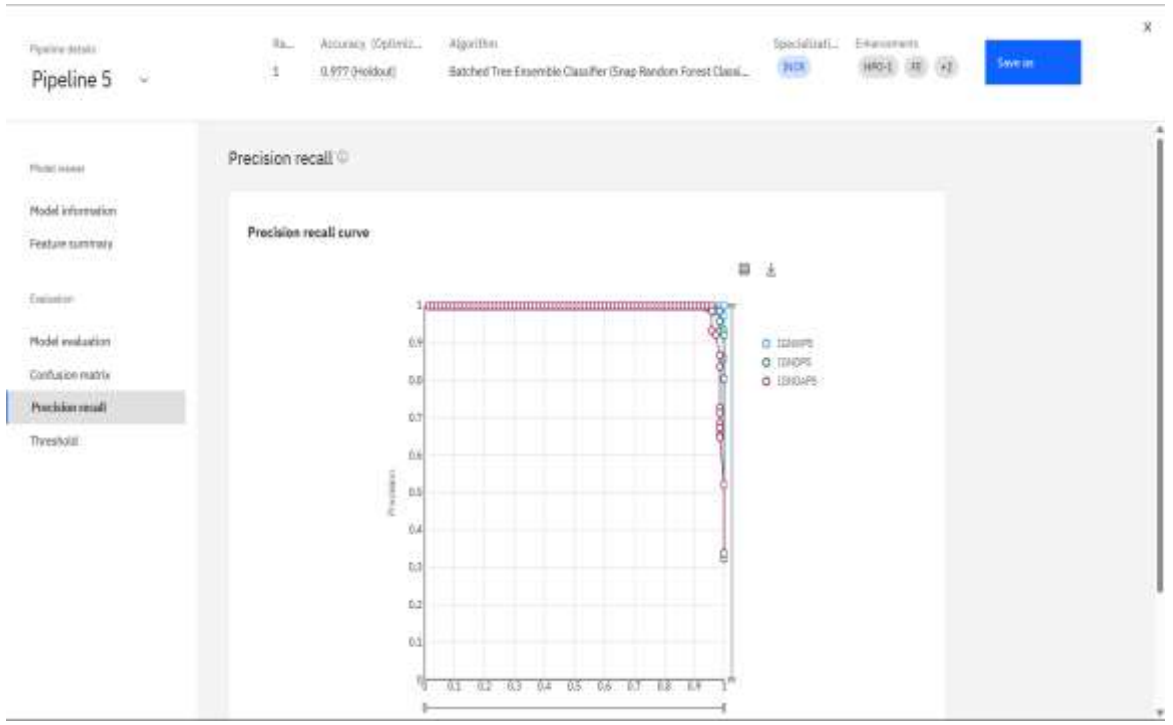
- Hosted on IBM Cloud Lite with built-in monitoring of latency and accuracy
- Periodic retraining ensures model freshness

RESULT

CONFUSION MATRIX



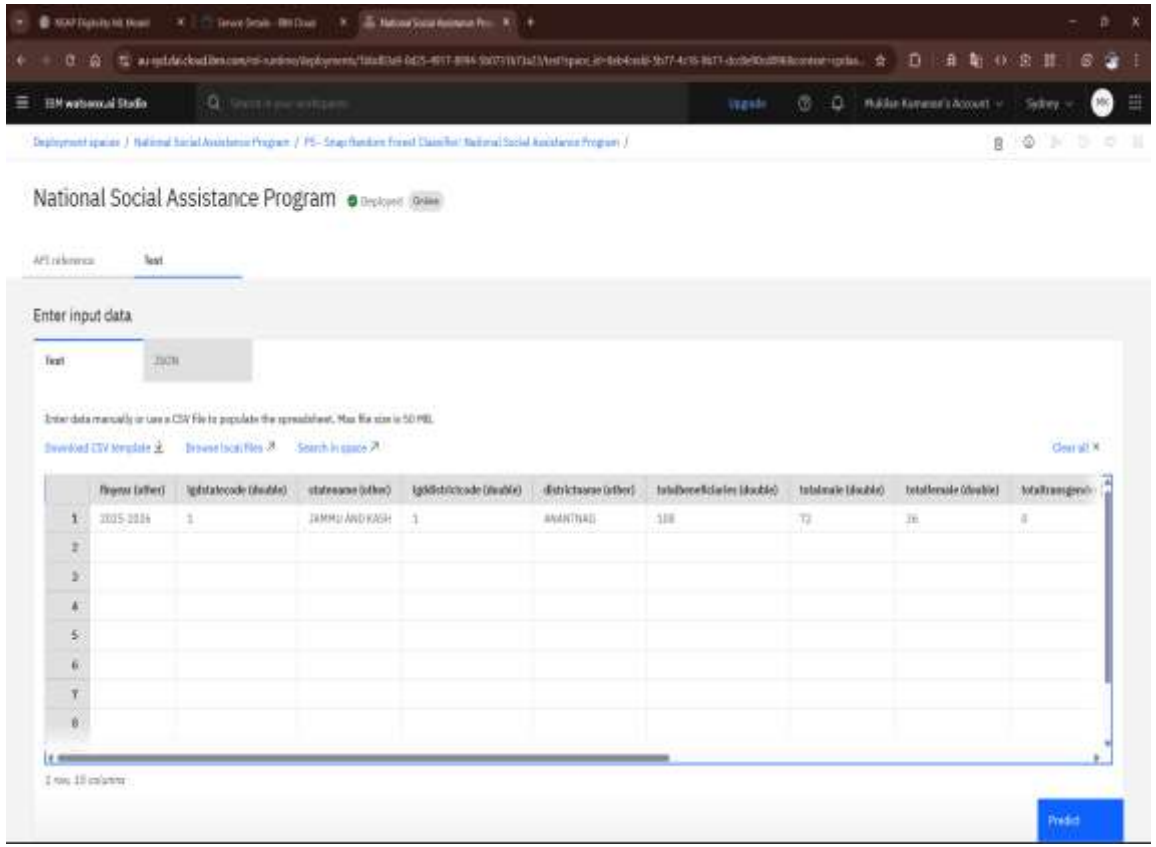
PRECISION RECALL

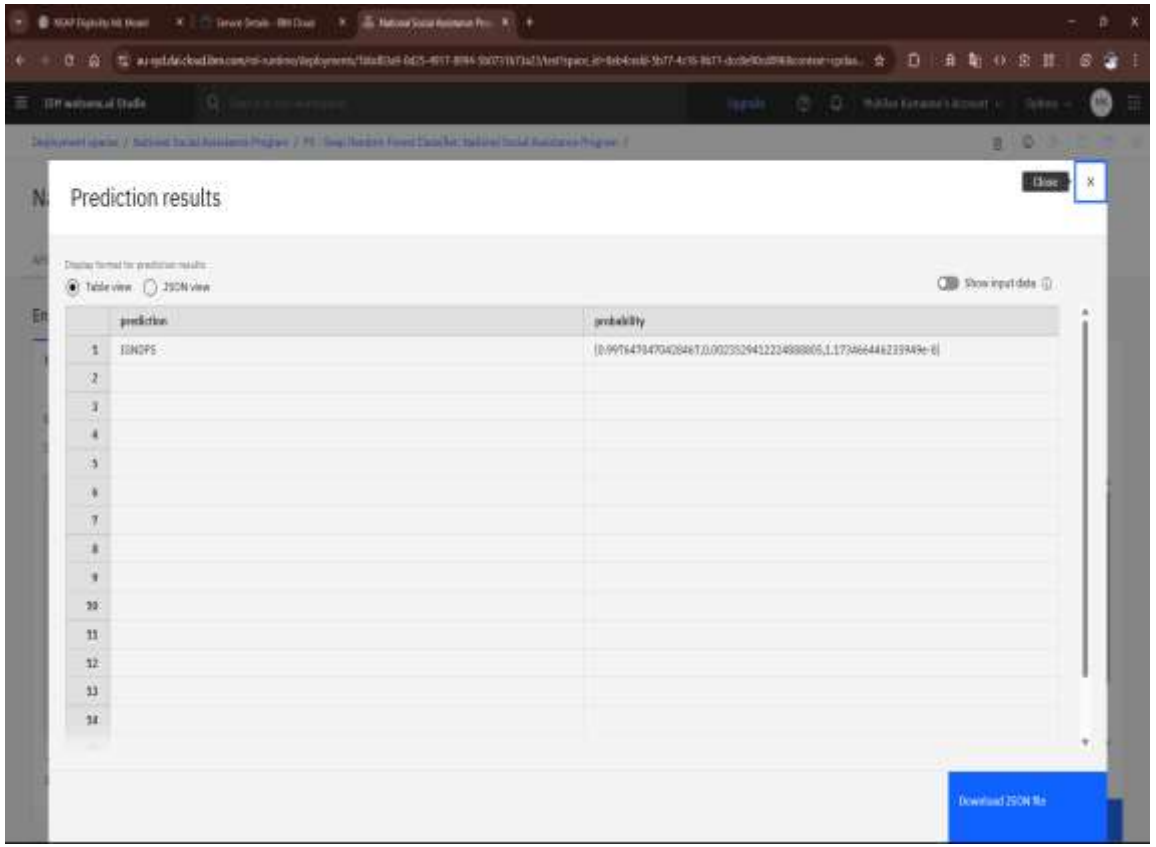


edunet foundation

RESULT

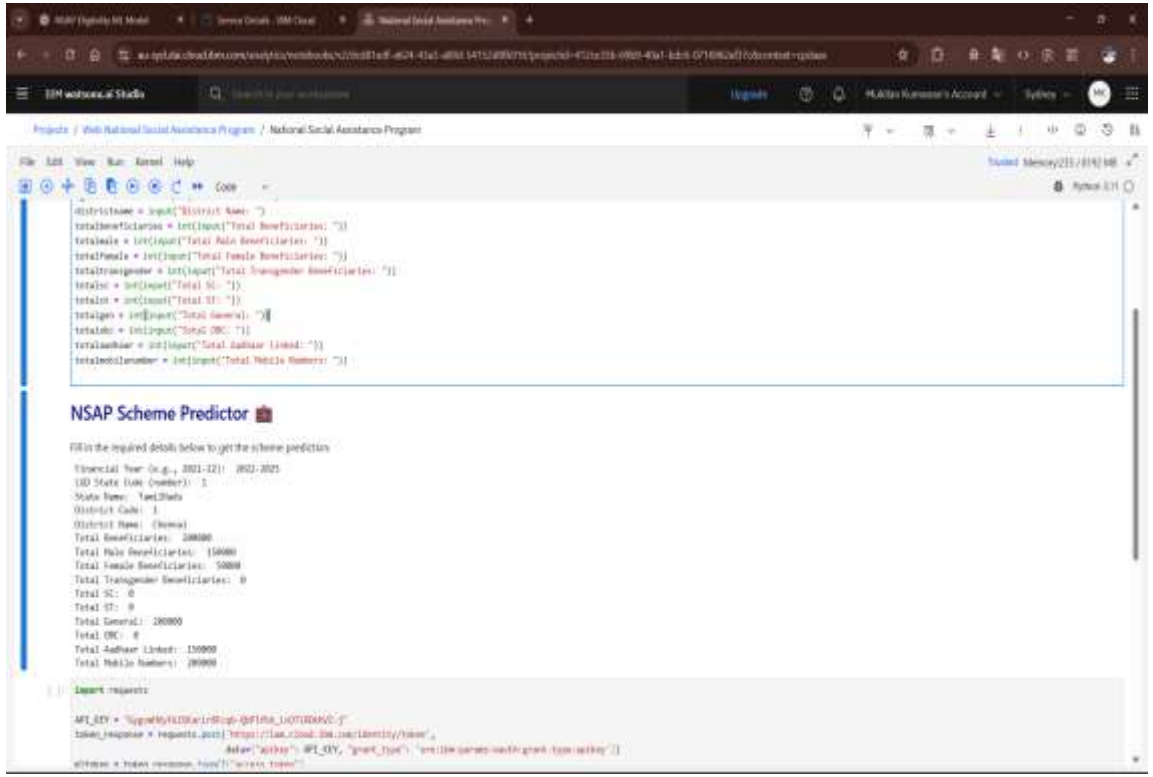
Model Preview and testing

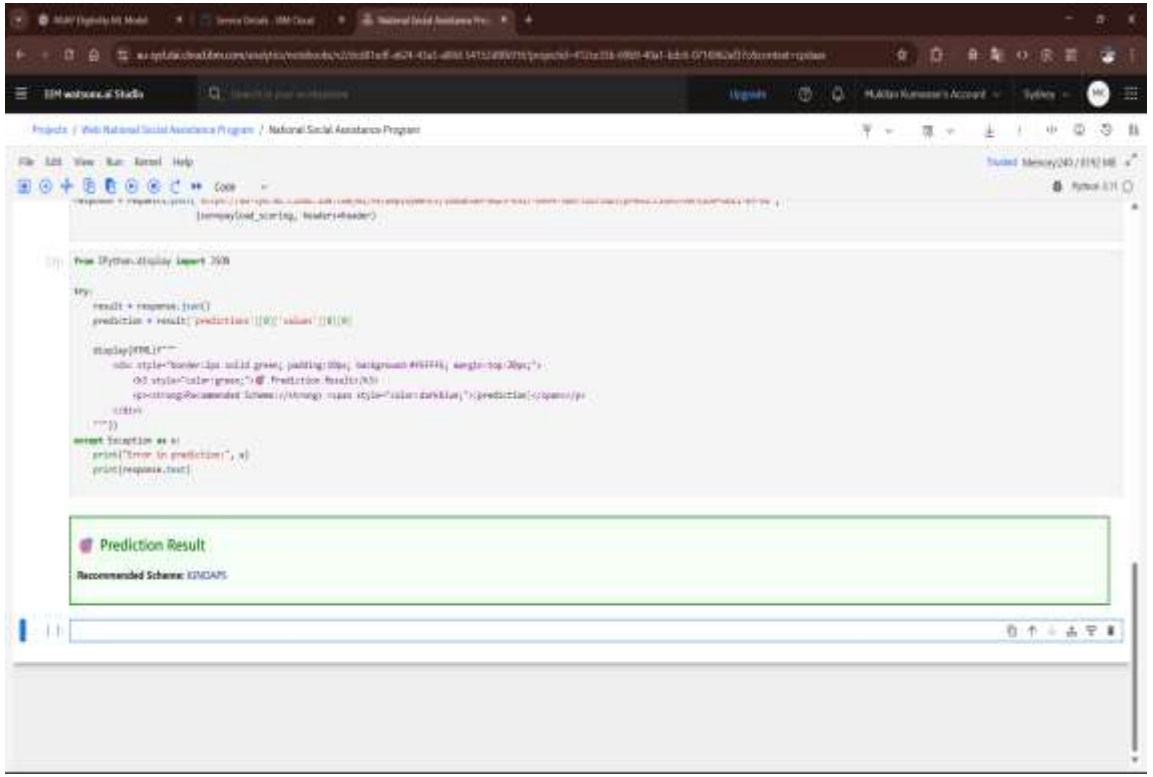




RESULT

Created A simple Script to run the ML model in the Notebook in IBM cloud





CONCLUSION

- The model effectively predicts the most suitable NSAP scheme for applicants using socio-economic and demographic data, reducing manual workload and errors.
- Deployment via IBM Watson Machine Learning ensures real-time access and seamless integration for government systems.
- Challenges faced included handling data imbalance and feature scaling. AutoAI helped streamline tuning and improved accuracy.
- Future improvements could include adding more regional or real-time applicant data to enhance prediction precision.
- Accurate scheme prediction ensures timely benefits reach deserving individuals, making the system more efficient, fair, and transparent.

FUTURE SCOPE

- **Integration of More Data Sources:** Future versions can incorporate real-time data from Aadhaar, local surveys, or health records to further enhance scheme prediction accuracy.
- **Algorithm Optimization:** The current Random Forest model can be enhanced using ensemble methods or deep learning models for better adaptability across diverse populations.
- **Scalability Across Regions**: The system can be expanded to other states or districts, supporting localized scheme recommendations tailored to regional policies.
- **Smart Integration:** Introducing edge AI devices in rural service centers could allow offline predictions and reduce dependency on internet connectivity, improving accessibility.

REFERENCES

- AI Kosh Dataset
  District-wise Pension Data – AI Kosh, IndiaAI
- IBM AutoAI Documentation
  AutoAI Guide – IBM Cloud Docs
- Random Forest Algorithm
  Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
- Model Evaluation Techniques
  Powers, D. (2011). Evaluation: Precision, Recall, F1 Score and ROC. arXiv:2010.16061.

GITHUB LINK

- GitHub Link : https://github.com/MUKILAN-K/NSAP-Prediction-IBMCloud-Edunet.git

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

MUKILAN K

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

edunet foundation

THANK YOU

edunet foundation