
CAPSTONE PROJECT

PREDICTING ELIGIBILITY FOR GOVERNMENT WELFARE SCHEMES USING MACHINE LEARNING

Presented By:

**Swati Upadhyay - Vaugh Institute of Agriculture Engineering and
Technology - BCA Department**

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Result (Output Image)
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

- India's National Social Assistance Programme (NSAP) provides essential financial support to vulnerable groups such as senior citizens, widows, and persons with disabilities.
- Currently, the process of verifying beneficiary eligibility is **manual, time-consuming, and prone to errors** especially when implemented at scale across diverse populations.
- The key challenge is to **automate the identification process** by predicting the most appropriate NSAP welfare scheme for an individual based on their **demographic and socio-economic features**.

PROPOSED SOLUTION

- The proposed system aims to address the challenge of automating eligibility prediction for NSAP welfare schemes using machine learning. The objective is to classify individuals into the most appropriate welfare scheme (IGNOAPS, IGNWPS, or IGNDPS) based on their demographic and socio-economic attributes. The solution consists of the following components:
- **Data Collection:**
 - Utilized structured data containing district-wise details such as gender distribution, caste categories (SC/ST/OBC/GEN), Aadhaar and mobile linkage, etc.
 - Data was sourced from NSAP public records and uploaded to IBM Cloud for modeling.
- **Data Preprocessing:**
 - Cleaned and validated the dataset to remove inconsistencies and ensure integrity.
 - Performed exploratory data analysis and feature extraction for relevant attributes influencing scheme assignment.
- **Machine Learning Algorithm:**
 - Implemented a supervised classification model (Random Forest)
 - Split data into training and testing sets
 - Trained model to predict NSAP scheme eligibility based on user profile
- **Deployment:**
 - Built and executed the solution in IBM Watson Studio (Cloud) using Python
- **Evaluation:**
 - Measured model performance using Accuracy, Confusion Matrix, and Classification Report
 - Achieved an accuracy of 97.2%, indicating strong generalization.
 - Result: The model successfully classifies each record into the appropriate NSAP welfare scheme with high confidence.

SYSTEM APPROACH

The “System Approach” outlines the overall strategy and methodology for developing and implementing the NSAP eligibility prediction system using machine learning on IBM Cloud.

■ **System Requirements :**

- IBM Cloud account with Cloud Object Storage and Watson Studio
- Jupyter Notebook environment (Python 3.11)
- Uploaded NSAP dataset in .csv format
- GitHub for version control and final submission
- Browser and stable internet connection

■ **Libraries Required to Build the Model :**

- pandas – for data handling and preprocessing
- numpy – for numerical operations
- matplotlib / seaborn – for data visualization
- scikit-learn – for ML model creation, training, and evaluation
- joblib – for model serialization (optional)
- warnings – to suppress runtime warnings
- IPython.display – for rendering output (optional)

ALGORITHM & DEPLOYMENT

■ Algorithm Selection:

- A Random Forest Classifier, a supervised machine learning algorithm, was selected due to its robustness, high accuracy, and ability to handle both numerical and categorical data.
- It performs well on classification tasks involving tabular, structured data making it ideal for predicting NSAP welfare schemes based on demographic attributes.

■ Data Input:

- The model was trained on district-wise aggregated data with the following input features:
 - totalmale, totalfemale, totaltransgender
 - totalsc, totalst, totalgen, totalobc
 - totalaadhaar, totalmobilenumber
- The target variable was: schemecode (IGNOAPS, IGNWPS, IGNDPS)

■ Training Process:

- The dataset was split into training and testing sets using an 80:20 ratio.
- The Random Forest Classifier was trained on the training set using default hyperparameters.
- Performance was validated using classification metrics such as accuracy, precision, recall, and confusion matrix.

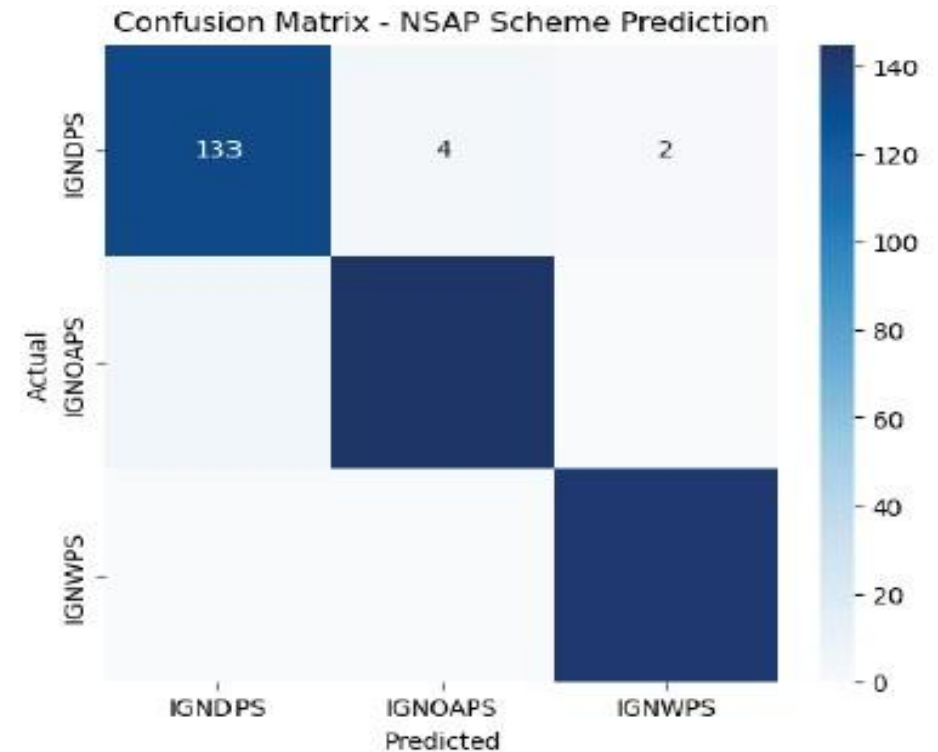
■ Prediction Process:

- The trained model was used to predict the appropriate welfare scheme for each record in the test set.
- Predictions were made entirely within IBM Cloud's Jupyter Notebook environment without real-time data input, as the dataset was pre-aggregated.
- The model achieved an accuracy of 97.2%, making it a reliable tool for automated scheme classification.

RESULT

The trained machine learning model was evaluated using a test dataset to assess its ability to predict NSAP welfare schemes accurately. The following evaluation metrics and confusion matrix highlight its strong performance.

Classification Report:				
	precision	recall	f1-score	support
IGNDPS	0.96	0.96	0.96	139
IGNOAPS	0.97	0.96	0.97	151
IGNWPS	0.99	1.00	0.99	142
accuracy			0.97	432
macro avg	0.97	0.97	0.97	432
weighted avg	0.97	0.97	0.97	432



CONCLUSION

- The project successfully demonstrated the potential of machine learning in automating the eligibility prediction process for government welfare schemes under the NSAP initiative. Using demographic and socio-economic features, the model was able to predict scheme categories (IGNOAPS, IGNWPS, IGNDPS) with **97.2% accuracy**.
- Throughout the project, careful preprocessing and feature analysis ensured data quality, while the Random Forest classifier provided robust results with minimal overfitting.
- Some challenges included:
 - Understanding data imbalance across scheme categories.
 - Handling subtle feature overlaps between eligible groups.
- These can be addressed further with advanced techniques like SMOTE or deep learning models.
- Overall, this approach can significantly reduce the manual effort and time in screening potential beneficiaries, enabling faster and more inclusive delivery of social support.
- Complete project code available at: <https://github.com/SwatiUpadhyay/NSAP-Eligibility-Prediction-IBM-Project>

FUTURE SCOPE

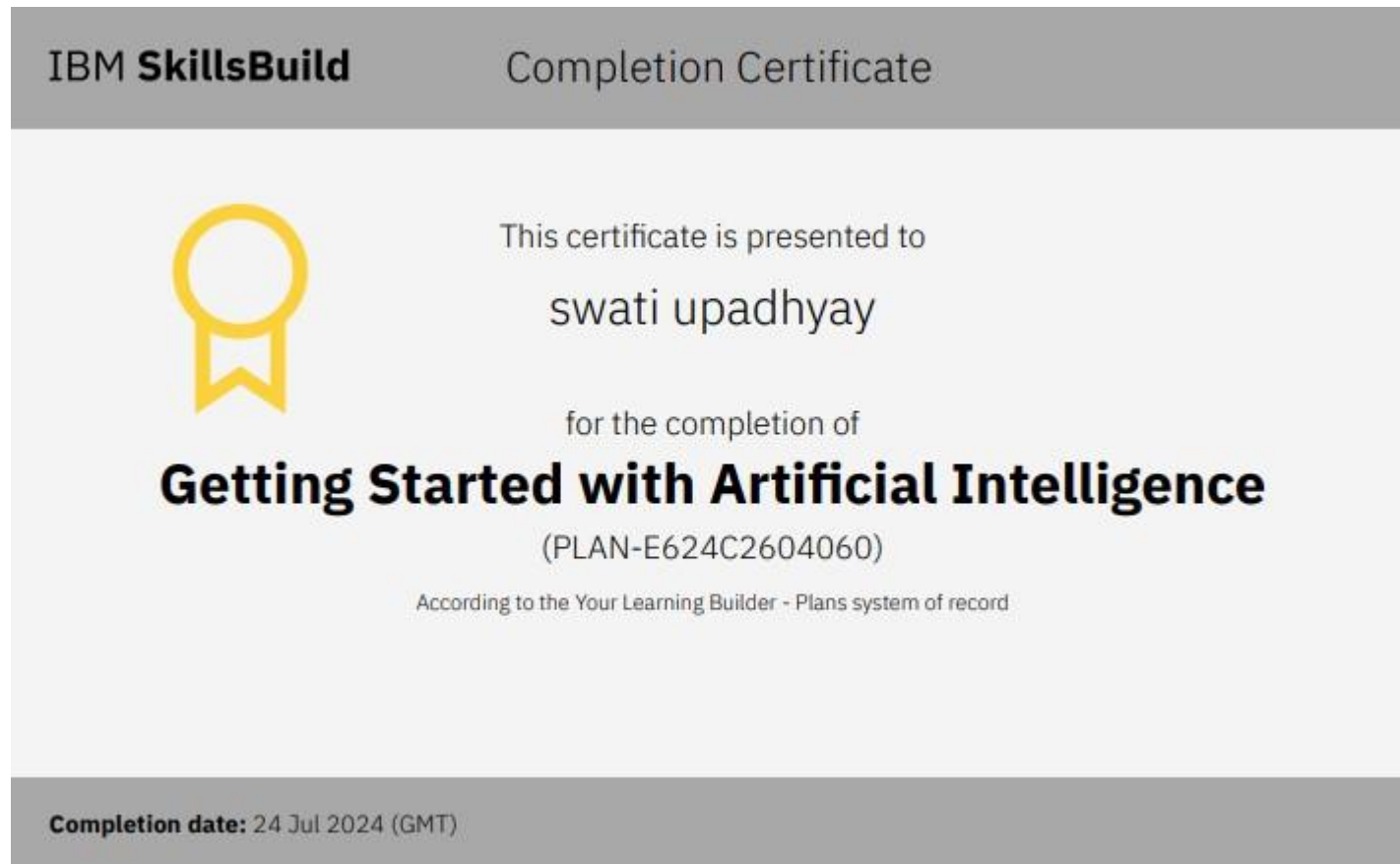
- To enhance the effectiveness and scalability of the system, several future improvements are proposed:
- **Incorporation of Additional Data Sources:**
 - Integrate more demographic, health, or economic indicators to refine prediction accuracy.
- **Algorithm Optimization:**
 - Experiment with ensemble models, gradient boosting, or neural networks (e.g., XGBoost, LightGBM) to improve performance.
- **Geographical Expansion:**
 - Extend the system to cover all Indian states and union territories, enabling a unified national framework.
- **Real-Time Integration:**
 - Deploy as a live web service that can accept user input and provide instant scheme recommendations.
- **Edge & Cloud Computing Integration:**
 - Explore deployment on edge devices for rural reach, and use IBM Cloud services to scale access.
- **Explainable AI (XAI):**
 - Incorporate interpretability methods (e.g., SHAP, LIME) to explain why a scheme is recommended for transparency.

REFERENCES

- Government of India - Ministry of Rural Development. *National Social Assistance Programme (NSAP)*: <https://nsap.nic.in>
- IBM SkillsBuild & Edunet Foundation Internship Materials: <https://sb4a.edunetworld.com>
- scikit-learn: Machine Learning in Python Pedregosa et al., *Journal of Machine Learning Research*, 2011: <https://scikit-learn.org>
- pandas: Python Data Analysis Library: <https://pandas.pydata.org>
- Kaggle Dataset - NSAP Welfare Scheme Records (*If applicable for citation*)
- IBM Cloud Documentation-Watson Studio, Cloud Object Storage: <https://cloud.ibm.com/docs>
- Complete project code available at: <https://github.com/SwatiUpadhyay/NSAP-Eligibility-Prediction-IBM-Project>

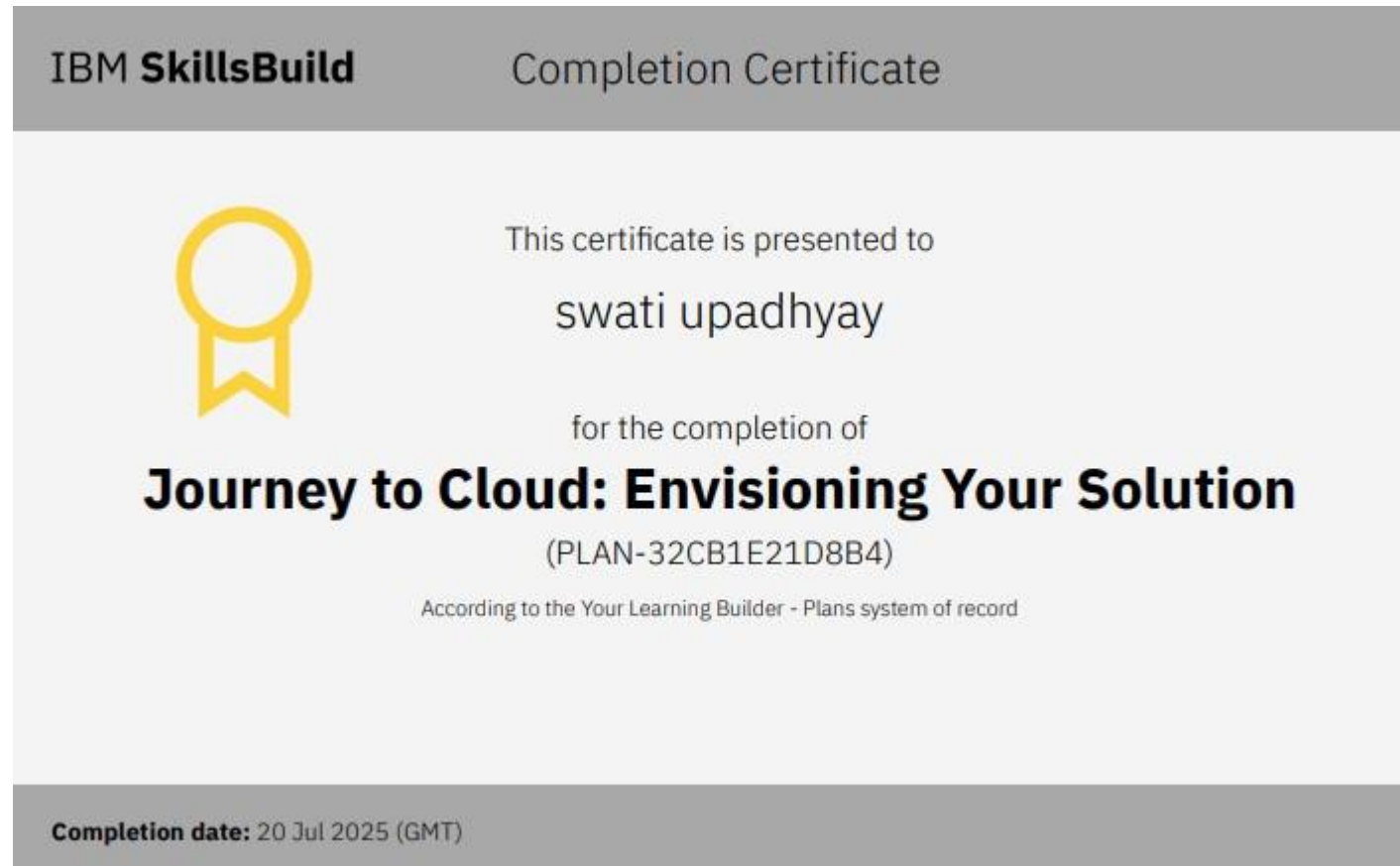
IBM CERTIFICATIONS

- Screenshot/ credly certificate(getting started with AI)



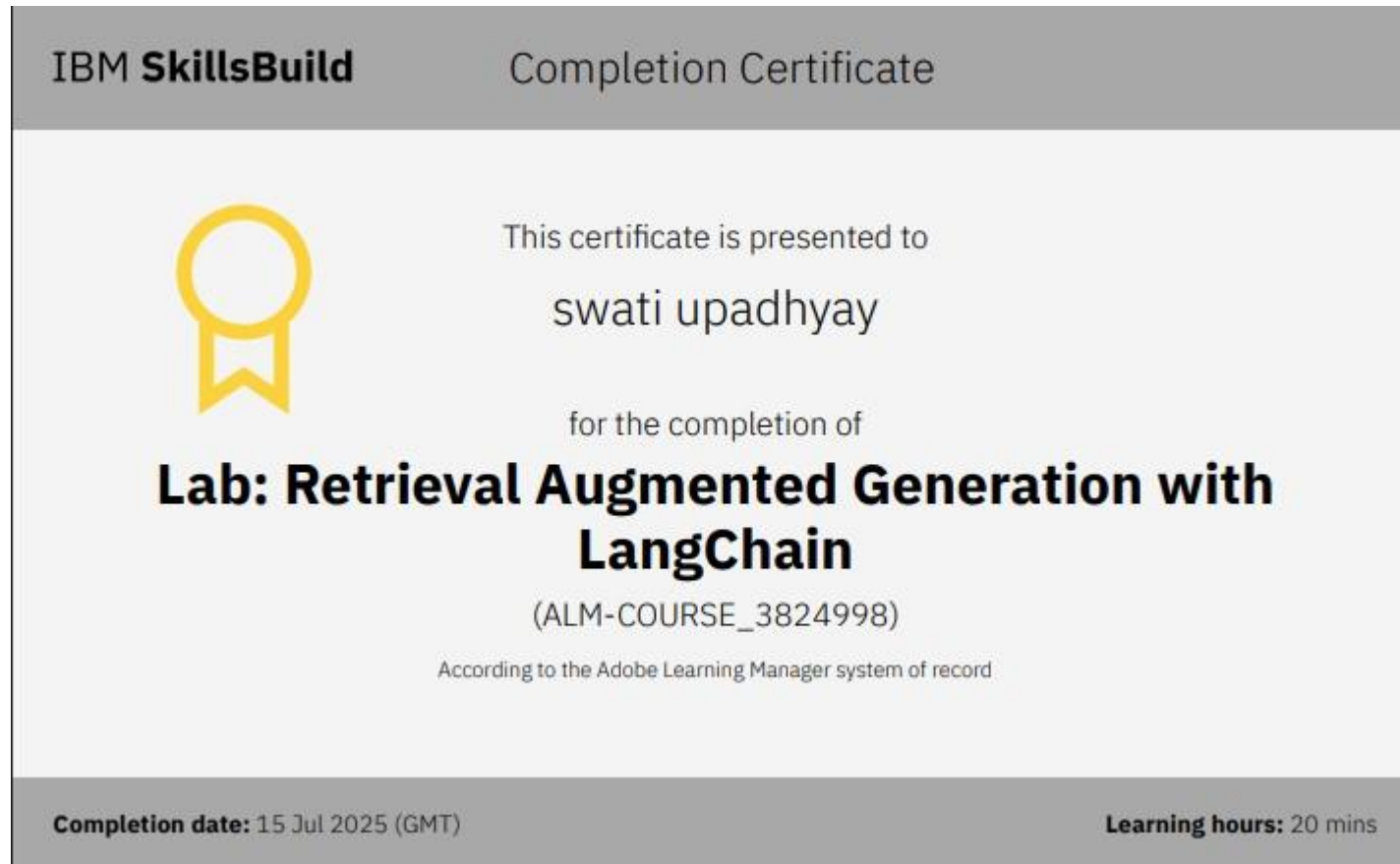
IBM CERTIFICATIONS

- Screenshot/ credly certificate(Journey to Cloud)



IBM CERTIFICATIONS

- Screenshot/ credly certificate(RAG Lab)





THANK YOU