
CAPSTONE PROJECT

PREDICTING PENSION ELIGIBILITY UNDER NSAP

Presented By:
Banka Spandana Priya
Kakatiya Institute of Technology and Science
Electrical and Electronics Engineering

OUTLINE

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach**
- **Algorithm & Deployment**
- **Result**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

Right now, figuring out who qualifies for pensions under the National Social Assistance Programme (NSAP) is quite a tough nut to crack. This is mainly because of the wide range and sheer amount of demographic and socio-economic data that varies from one district to another. We really need a solid approach to automate the process and enhance the accuracy of pension eligibility decisions. This way, we can ensure fairness and make sure that government resources are targeted efficiently.

PROPOSED SOLUTION

- In this project, I created a system that predicts pension eligibility under the National Social Assistance Programme (NSAP) using machine learning techniques.
- I gathered district-level data on demographics, socio-economic factors, and Aadhaar/mobile coverage relevant to pension eligibility.
- The data was cleaned and preprocessed to address missing values, outliers, and inconsistencies. I also developed meaningful features to help the model learn.
- I trained a Random Forest Classifier on the prepared dataset to distinguish accurately between eligible and non-eligible beneficiaries.
- Finally, I deployed the trained model on IBM Watson Machine Learning, allowing for scalable, real-time pension predictions through REST API.

SYSTEM APPROACH

For this project, I used the IBM watsonx.ai Studio environment and the Watson Machine Learning service instance from IBM Cloud. To safely store and manage our input datasets, I used IBM Cloud Object Storage.

Library required to build the model

- To create the predictive model, I worked a lot with these Python libraries:
- pandas for data manipulation and preprocessing
- numpy to handle numerical computations
- scikit-learn to develop and train the Random Forest Classifier model
- matplotlib and seaborn (optional) for data visualization and analysis
- ibm-watson-machine-learning SDK for managing, storing, and deploying the trained models on the cloud platform

ALGORITHM & DEPLOYMENT

Algorithm Selection:

- I chose the Random Forest Classifier to predict pension eligibility. This algorithm works well because it effectively handles diverse demographic and socio-economic data, manages missing values, and reduces overfitting through ensemble learning. Its interpretability and strong performance for classification tasks support this choice given the varied and complex input data.

Data Input:

- The model uses input features, including district-level demographic counts (e.g., total male, total female, SC/ST/OBC populations), Aadhaar and mobile phone coverage statistics, and other socio-economic indicators taken from NSAP data.

Training Process:

- The model was trained on cleaned and preprocessed historical pension eligibility data. I used cross-validation techniques to ensure robust performance estimation and adjusted hyperparameters such as the number of trees and maximum depth to improve accuracy.

Prediction Process:

- After training, the model predicts pension eligibility by analyzing new input records in real-time once deployed. It provides class probabilities and predicted labels through the IBM Watson Machine Learning RESTful API. The prediction system can include updated data sources to maintain accuracy over time.

RESULT

- In my project, the Random Forest Classifier model performed well in predicting pension eligibility under the NSAP. The model showed high accuracy and reliability when tested on new district-level data.
- The prediction output includes both the eligibility class and the related confidence probabilities, which helps in making informed decisions.
- Visualizations that compare predicted eligibility with actual data confirm that the model effectively identifies eligible beneficiaries and supports targeted resource allocation.
- The results show that machine learning can significantly improve the efficiency and fairness of pension payments under NSAP.

CONCLUSION

- In this project, I developed and deployed a machine learning model to predict pension eligibility under the National Social Assistance Programme (NSAP). The model's strong predictive performance shows how data-driven methods can improve social welfare targeting.
- By automating eligibility assessment, this solution improves accuracy, ensures fairness, and optimizes resource use. Deploying the model on IBM Watson Machine Learning provides scalable, real-time access to predictions, making it practical for government use.
- This project shows the value of combining demographic data and analytics to support evidence-based policymaking and efficient distribution of social assistance.

FUTURE SCOPE

- Going forward, I plan to improve this pension eligibility prediction system by adding real-time data sources, such as updated demographic surveys and Aadhaar or mobile status changes, to increase prediction accuracy and responsiveness.
- I also aim to create user-friendly dashboards and interfaces that help government officials and stakeholders visualize prediction results and make informed decisions quickly.
- Exploring machine learning techniques, like deep learning, and including methods for explainability could further boost model performance and transparency.
- Additionally, this approach can be applied to other social welfare schemes and regions to expand its impact on public policy and social assistance programs.

REFERENCES

- Official reports and datasets from the National Social Assistance Programme (NSAP)
- Documentation and tutorials from IBM Watson Machine Learning, which guided the deployment and management of the model
- Academic research papers and articles focused on applying machine learning techniques to social welfare and eligibility prediction
- The scikit-learn library documentation for implementing and tuning the Random Forest classifier
- IBM Cloud Object Storage documentation to securely manage and store input datasets
- Additional scholarly articles on Random Forest algorithms and classification methods to support model selection and evaluation

IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Spandana Priya Banka

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/9d0772df-7077-4ecf-8aff-8bf3098e6233>



IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence



Spandana Priya Banka

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution



Issued on: Jul 18, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/8949ad1c-ff58-448c-a9f7-983d3f14c40f>



IBM CERTIFICATIONS

IBM **SkillsBuild**

Completion Certificate



This certificate is presented to
Spandana Priya Banka

for the completion of
**Lab: Retrieval Augmented Generation with
LangChain**

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 18 Jul 2025 (GMT)

Learning hours: 20 mins



THANK YOU