IBM AICTE PROJECT

PREDICTING ELIGIBILITY

Presented By:
1. Allu Likhith-Chandigarh University-Computer 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

Example: 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.

PROPOSED SOLUTION

- The proposed system aims to address the challenge of predicting whether an individual is eligible for the National Social Assistance Programme (NSAP) based on socio-economic and demographic data. This involves leveraging data preprocessing, feature engineering, and machine learning—specifically Ridge Regression—to build a robust predictive model. The solution will consist of the following components:
- **Data Collection:**
  - Use the provided NSAP dataset containing attributes such as age, gender, income, employment status, education level, and other socio-economic indicators.
  - Ensure all relevant features contributing to eligibility determination are included.
- **Data Preprocessing:**
  - Handle missing values, outliers, and inconsistencies in the dataset.
  - Apply encoding for categorical variables.
  - Normalize/scale numerical features to improve model convergence and performance.
  - Perform feature selection to remove irrelevant or redundant attributes.
- **Machine Learning Algorithm:**
  - Implement **Ridge Regression** to address multicollinearity and reduce overfitting.
  - Train the model using historical NSAP eligibility data.
  - Optimize the regularization parameter (alpha) using **GridSearchCV** or cross-validation.

edunet foundation

PROPOSED SOLUTION

- **Deployment:**
  - Build a user-friendly interface or API that accepts input attributes and returns eligibility predictions.
  - Deploy the model on a scalable cloud platform for accessibility.
- **Evaluation:**
  - Evaluate the model using metrics such as **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, and **R^2 score**.
  - Compare Ridge Regression performance with other models (if tested) to validate its effectiveness.
  - Continuously monitor and retrain the model as new data becomes available.
- **Expected Result**
  - An accurate, interpretable, and efficient system that can predict NSAP eligibility with minimal error.
  - A deployed application or notebook-based tool that aids government officials and policymakers in decision-making.

edunet foundation

SYSTEM APPROACH

The system approach for predicting eligibility for the National Social Assistance Programme (NSAP) focuses on the step-by-step methodology to design, develop, and deploy a machine learning-based predictive model. This approach ensures that the system is accurate, efficient, and scalable for real-world applications:

- **System requirements**
  - The system requires the following hardware and software components:
  - **Hardware Requirements**
  - **Processor:** Intel i5 or above
  - **RAM:** Minimum 8 GB
  - **Storage:** Minimum 500 GB HDD or 256 GB SSD
  - **GPU (optional):** For faster computation during training
  - **Software Requirements**
  - **Operating System:** Windows / Linux / macOS
  - **Operating System:** Windows / Linux / macOS
  - IDE / Notebook: Jupyter Notebook, VS Code, or PyCharm
  - Version Control: Git (optional, for collaborative work)

edunet foundation

SYSTEM APPROACH

- Library required to build the model
  - Handle missing values, outliers, and inconsistencies in the dataset.
  - Apply encoding for categorical variables.
  - Pandas-> Data loading, cleaning, and preprocessing.
  - Numpy-> Numerical computations.
  - Matplotlib-> Data Visualization.
  - seaborn-> Statistical Data visualization
  - Scikit-learn-> Model training, Ridge Re

edunet foundation

ALGORITHM & DEPLOYMENT

- Algorithm Selection:
  - For predicting eligibility, the chosen machine learning algorithm is **Ridge Regression**, a regularized linear regression technique. Ridge Regression introduces an L2 penalty term to the loss function, which helps to prevent overfitting by shrinking large coefficients. This makes it particularly effective when dealing with datasets that may have multicollinearity or a large number of features with varying scales.
    The algorithm was selected because:
  - The problem involves predicting a **continuous score/label** related to eligibility, making regression models a natural fit.
  - Ridge Regression balances model complexity and generalization through its regularization parameter.
  - The dataset size and structure make linear models computationally efficient while still delivering accurate results.
- Data Input:
  - The model takes in a set of structured features from the eligibility dataset. These include:
  - **Demographic variables** (e.g., age, gender, region)
  - **Socio-economic indicators** (e.g., income level, employment status)
  - **Program-specific attributes** (e.g., household size, disability status, scheme participation history)
  - **Other relevant categorical or numerical fields** extracted from the original dataset.

edunet
foundation

ALGORITHM & DEPLOYMENT

- **Training Process:**
  - The training process was executed using **IBM Watsonx AutoAI**:
  - **Data Connection Setup:** Data was linked from IBM Cloud Object Storage using the DataConnection API.
  - **Automated Preprocessing:** AutoAI automatically handled missing values, feature encoding, and scaling.
  - **Model Search & Optimization:** AutoAI explored multiple algorithms but selected Ridge Regression as the best performer based performer based on evaluation metrics.
  - **Cross Validation:** The model was validated using k-fold cross-validation to assess generalization performance.
  - **Hyperparameter Tuning:** The regularization strength (alpha) as tuned automatically to optimize prediction accuracy while avoiding overfitting.
- **Prediction Process:**
  - The final trained model was deployed on IBM Watson Machine Learning (WML) service.
  - **Deployment Type:** Online deployment for real-time scoring.
  - **API Access:** WML provides a REST API endpoint to send applicant data and receive eligibility predictions.
  - **Scalability:** The deployment can handle multiple concurrent requests, making it suitable for integration into large-scale eligibility screening platforms.
  - **Monitoring: I**BM Watsonx provides monitoring tools to track prediction accuracy, latency, and data drift over time.

RESULT

The Ridge Regression model demonstrated strong performance in predicting eligibility scores, as evaluated by IBM Watsonx AutoAI. During model selection, AutoAI compared multiple algorithms and selected Ridge Regression for its superior balance of accuracy and generalization.

- **Performance Metrics**
  - While AutoAI handles much of the evaluation internally, the chosen Ridge model achieved:
  - **High R^2 score** – indicating that the model explains a significant portion of the variance in the target variable.
  - **Low RMSE (Root Mean Squared Error)** – showing minimal deviation between predicted and actual eligibility scores.
  - **Stable cross-validation results** – suggesting consistent performance across different subsets of the data.
  - **Effectiveness**
  - The model's predictive strength means it can be confidently deployed for real-time eligibility assessment. Its regularization reduces overfitting, making it robust to variations in input data.

CONCLUSION

The developed system using **Ridge Regression** within IBM Watsonx AutoAI has proven effective in predicting eligibility with high accuracy and reliability. By leveraging automated preprocessing, feature engineering, and hyperparameter optimization, the model delivers predictions that closely align with actual outcomes, making it suitable for real-time decision-making.

During implementation, the main challenges included:

- **Data integration** – Ensuring consistency in format and quality across input datasets.
- **Feature relevance** – Identifying and retaining only the most impactful predictors to avoid overfitting.
- **Platform-specific dependencies** – Managing IBM Watsonx AutoAI and WML service configurations to ensure smooth model deployment.

Despite these challenges, the solution successfully addresses the core objective: enabling automated, scalable, and accurate eligibility assessment. Future enhancements, such as incorporating additional datasets, adopting advanced algorithms, and expanding to multi-region deployments, will further strengthen its performance.

Accurate prediction of eligibility is critical for ensuring that benefits and resources are allocated efficiently and fairly. A robust, automated system not only accelerates processing but also improves transparency and trust in the decision-making process, ultimately supporting equitable service delivery at scale.

FUTURE SCOPE

The current system for predicting eligibility using Ridge Regression demonstrates high accuracy and practical applicability. However, there is significant potential to enhance and expand the system for greater coverage, robustness, and adaptability.

1. Incorporating Additional Data Sources

•**Socio-economic data integration** – Include richer datasets such as census records, healthcare access data, or community-level indicators to improve prediction accuracy.

•**Behavioral and transactional data** – Integrate applicant transaction histories, utility bill patterns, or mobility data for deeper insights.

•**Real-time data streams** – Use live updates from government databases to provide more timely and relevant predictions.

2. Algorithm Optimization

•**Advanced regularization techniques** – Experiment with Elastic Net or Bayesian Regression to combine the benefits of L1 and L2 penalties for improved feature selection and robustness.

•**Ensemble methods** – Explore Gradient Boosting, Random Forests, or hybrid approaches to potentially outperform a single Ridge Regression model.

•**Hyperparameter optimization** – Implement Bayesian Optimization or Genetic Algorithms to fine-tune model parameters more effectively.

3. Scalability and Regional Expansion

•**Multi-city deployment** – Extend the model to handle data from different regions, adapting to local eligibility rules and socio-economic patterns.

•**Transfer learning approaches** – Retrain the model with small amounts of new city-specific data to enable rapid scaling across geographies.

4. Integration of Emerging Technologies

•**Edge Computing** – Deploy the model on low-power devices (e.g., kiosks, community centers) to enable offline eligibility checking in remote areas.

•**Federated Learning** – Train models across multiple institutions without sharing sensitive applicant data, ensuring privacy compliance.

•**Explainable AI (XAI)** – Integrate interpretability tools like SHAP or LIME to help decision-makers understand why an applicant was predicted as eligible or not.

5. Continuous Monitoring and Improvement

•**Data drift detection** – Continuously monitor changes in input data distributions and retrain the model when performance drops.

•**Feedback loop integration** – Use actual eligibility outcomes to refine and update the model for better accuracy over time.

REFERENCES

1.Hoerl, A. E., & Kennard, R. W. (1970). *Ridge Regression: Biased Estimation for Nonorthogonal Problems*. Technometrics, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634
2.Kuhn, M., & Johnson, K. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press.
3.Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.
4.IBM Documentation. (2025). *IBM Watsonx AutoAI Overview*. Retrieved from https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/autoai-overview.html
5.Zhang, Y., & Yang, Q. (2017). *A Survey on Multi-Task Learning*. arXiv preprint arXiv:1707.08114. https://arxiv.org/abs/1707.08114
6.Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
7.Raschka, S., & Mirjalili, V. (2022). *Machine Learning with PyTorch and Scikit-Learn*. Packt Publishing.
8.Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Getting Started with Artificial Intelligence

IBM SkillsBuild

LIKHITH ALLU

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 22, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/e27e92b2-668e-4d21-8258-98f9213df8f0

IBM

edunet foundation

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

LIKHITH ALLU

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Aug 07, 2025

Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/4f41a90e-8316-4525-8fd1-312f3c52435b

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

LIKHITH ALLU

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 07 Aug 2025 (GMT) **Learning hours:** 20 mins

GITHUB LINK

- Github Link: https://github.com/Likhithallu/Predicting-Eligibility

THANK YOU