CAPSTONE PROJECT

SALARY PREDICTOR

Presented By:
1. Student Name- GOURAB GORAI
2. COLLEGE NAME – DR. B. C. ROY ACADEMY OF PROFESSIONAL COURSES
3. AICTE ID :- STU65f3159ce3fe81710429596
4. DEPT. :- BCA

edunet
foundation

OUTLINE

- **Problem Statement** (Should not include solution)
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment (Step by Step Procedure)**
- **Result**
- **Conclusion**
- **Future Scope(Optonal)**
- **References**

PROBLEM STATEMENT

- **Problem Statement: Salary Prediction Web Application Using Machine Learning**
- Modern job seekers and HR professionals often struggle to estimate fair and accurate salaries based on various candidate attributes such as education, experience, location, and job title. This lack of transparency can lead to misaligned expectations and hiring inefficiencies. The goal of this project is to develop a **user-interactive Flask web application** that predicts a candidate’s salary using machine learning. The application must:
- Validate user inputs rigorously (e.g., age-experience consistency, missing fields).
- Train a **Random Forest Regression model** using historical salary data.
- Evaluate the model using **R^2 score**, **MSE**, and **visualizations** like residual plots and feature importance charts.
- Handle missing or malformed datasets gracefully.
- Display predictions along with graphical insights to enhance interpretability.
- The system serves as a robust decision support tool for individuals and organizations by ensuring reliable predictions and real-time model feedback through a web interface.

SYSTEM APPROACH

- System Requirements
- Hardware:
  - Processor: Intel i3 or higher
  - RAM: Minimum 4 GB
  - Storage: Minimum 100 MB for datasets and dependencies
- Software:
  - Python 3.7 or above
  - Web browser (for accessing Flask app locally)
  - Operating System: Windows/Linux/macOS

SYSTEM APPROACH

- Libraries and Frameworks Used
  - **Flask:** Web framework to build the interactive UI
  - **Pandas:** Data manipulation and preprocessing
  - **NumPy:** Numerical operations
  - **Matplotlib & Seaborn:** For generating evaluation plots
  - **scikit-learn (sklearn)**
- Implementation Strategy
- **Model Training:** On valid dataset using a Random Forest Regression pipeline
- **User Interface:** Accepts form inputs via Flask and validates them
- **Salary Prediction:** On-the-fly predictions using trained model
- **Evaluation:** Generates plots like Actual vs Predicted, Residuals, Feature Importance
- **Error Handling:** Displays user-friendly messages for invalid inputs or system errors

ALGORITHM & DEPLOYMENT

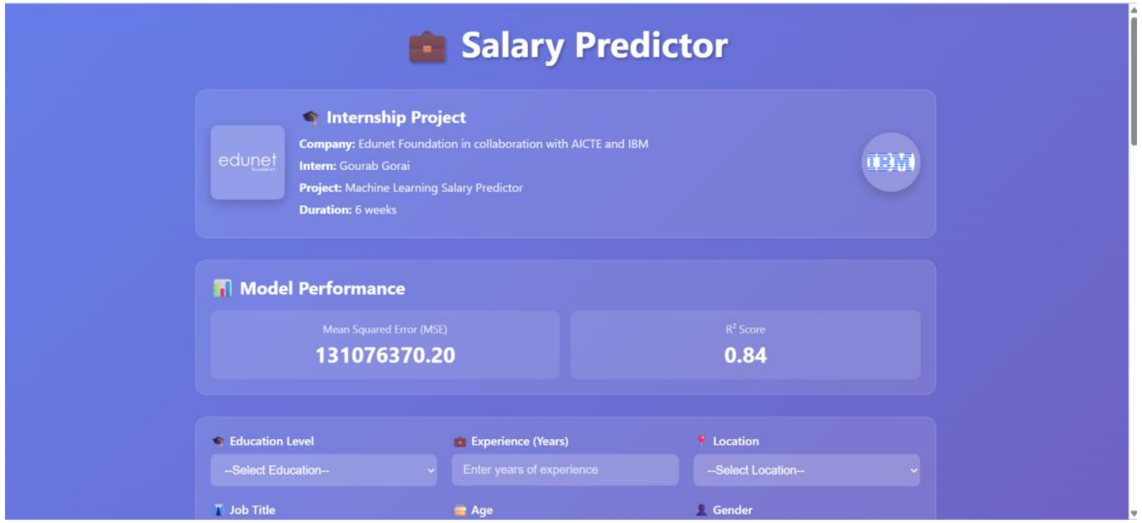
Step-by-Step Procedure:

- Data Collection
  - Load the dataset.
- Data Preprocessing
  - Check for missing or malformed entries.
  - Convert categorical features using OneHotEncoding.
  - Validate numerical fields
    - Age > 0 and < 100
    - Experience ≥ 0 and ≤ 70
    - Logical consistency between age and experience.

ALGORITHM & DEPLOYMENT

- Model Building
  - Split the dataset into training and testing sets using an 80:20 ratio.Use RandomForestRegressor wrapped inside a Pipeline that includes the preprocessing steps.
  - Train the model on the training data.
- Model Evaluation
  - Predict salaries on the test set.Evaluate the model using:R^2 Score
  - Mean Squared Error (MSE)
  - Generate the following plots for model analysis:
    - Line Plot: Actual vs Predicted Salaries
    - Histogram: Distribution of Residuals
    - Bar Chart: Feature Importance
    - Scatter Plot: Actual vs Predicted Salaries
- Web Application Integration
  - Build a web interface using the Flask framework. Allow users to enter details through an HTML form.Validate user inputs on the server side.Use the trained model to predict salary based on the user's inputs.Display the predicted salary and relevant evaluation plots.

RESULT



💼 Salary Predictor

🎓 Internship Project

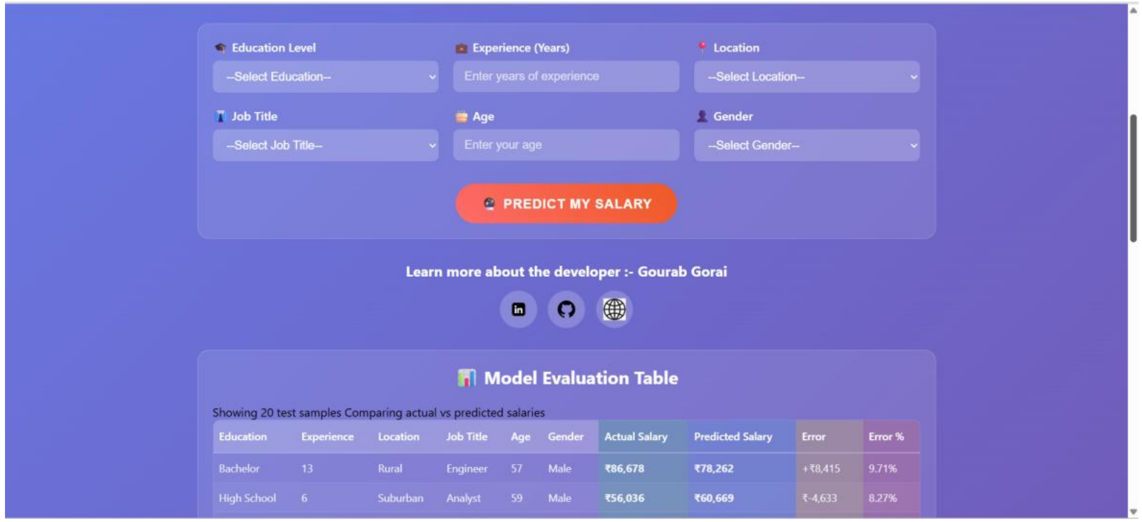
Company: Edunet Foundation in collaboration with AICTE and IBM
Intern: Gourab Gorai
Project: Machine Learning Salary Predictor
Duration: 6 weeks

📊 Model Performance

Mean Squared Error (MSE)
131076370.20

R² Score
0.84

🎓 Education Level: --Select Education--
💼 Experience (Years): Enter years of experience
📍 Location: --Select Location--
👔 Job Title
🎂 Age
👤 Gender



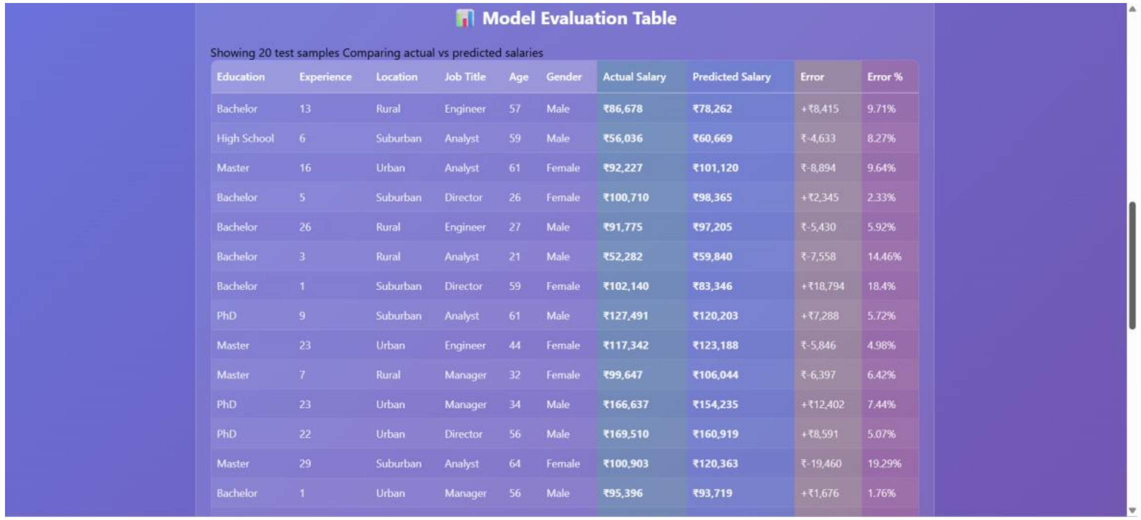
🎓 Education Level: --Select Education--
💼 Experience (Years): Enter years of experience
📍 Location: --Select Location--
👔 Job Title: --Select Job Title--
🎂 Age: Enter your age
👤 Gender: --Select Gender--

Learn more about the developer :- Gourab Gorai

📊 Model Evaluation Table

Showing 20 test samples Comparing actual vs predicted salaries

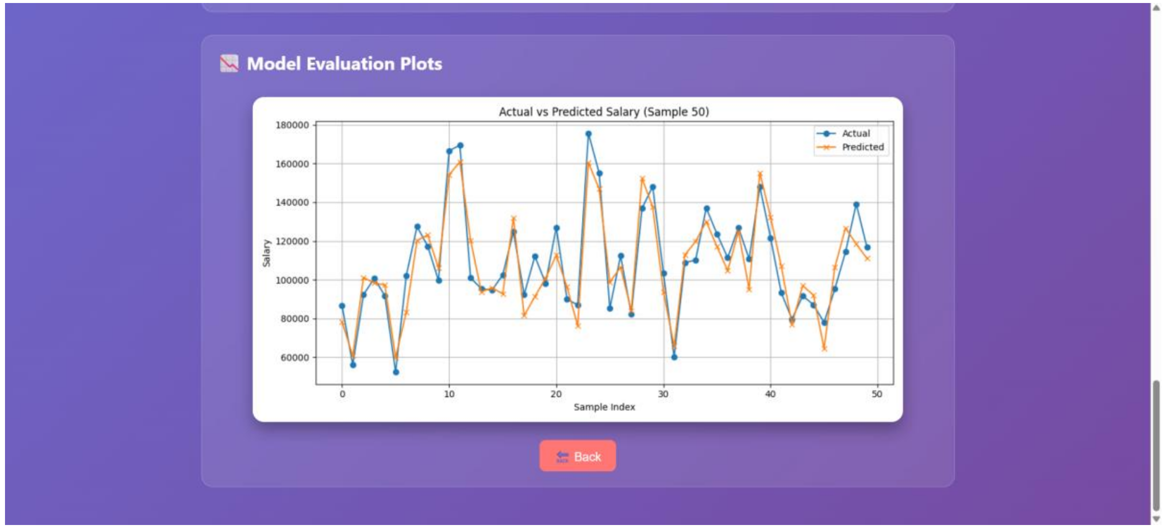
Education	Experience	Location	Job Title	Age	Gender	Actual Salary	Predicted Salary	Error	Error %
Bachelor	13	Rural	Engineer	57	Male	₹86,678	₹78,262	+₹8,415	9.71%
High School	6	Suburban	Analyst	59	Male	₹56,036	₹60,669	₹-4,633	8.27%



📊 Model Evaluation Table

Showing 20 test samples Comparing actual vs predicted salaries

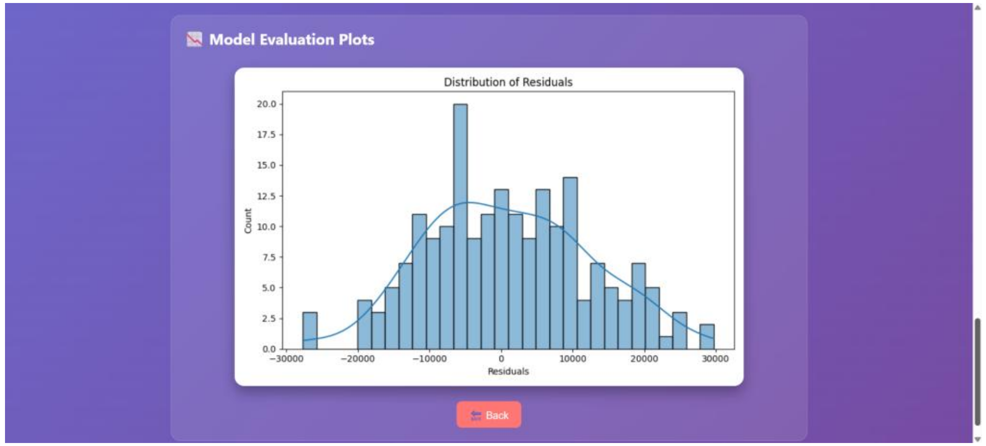
Education	Experience	Location	Job Title	Age	Gender	Actual Salary	Predicted Salary	Error	Error %
Bachelor	13	Rural	Engineer	57	Male	₹86,678	₹78,262	+₹8,415	9.71%
High School	6	Suburban	Analyst	59	Male	₹56,036	₹60,669	₹-4,633	8.27%
Master	16	Urban	Analyst	61	Female	₹92,227	₹101,120	₹-8,894	9.64%
Bachelor	5	Suburban	Director	26	Female	₹100,710	₹98,365	+₹2,345	2.33%
Bachelor	26	Rural	Engineer	27	Male	₹91,775	₹97,205	₹-5,430	5.92%
Bachelor	3	Rural	Analyst	21	Male	₹52,282	₹59,840	₹-7,558	14.46%
Bachelor	1	Suburban	Director	59	Female	₹102,140	₹83,346	+₹18,794	18.4%
PhD	9	Suburban	Analyst	61	Male	₹127,491	₹120,203	+₹7,288	5.72%
Master	23	Urban	Engineer	44	Female	₹117,342	₹123,188	₹-5,846	4.98%
Master	7	Rural	Manager	32	Female	₹99,647	₹106,044	₹-6,397	6.42%
PhD	23	Urban	Manager	34	Male	₹166,637	₹154,235	+₹12,402	7.44%
PhD	22	Urban	Director	56	Male	₹169,510	₹160,919	+₹8,591	5.07%
Master	29	Suburban	Analyst	64	Female	₹100,903	₹120,363	₹-19,460	19.29%
Bachelor	1	Urban	Manager	56	Male	₹95,396	₹93,719	+₹1,676	1.76%



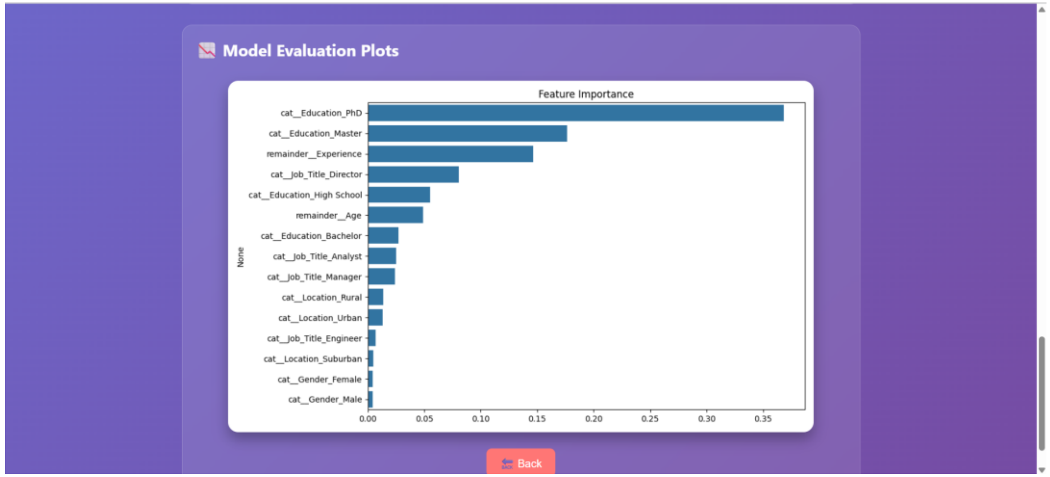
📉 Model Evaluation Plots

Actual vs Predicted Salary (Sample 50)

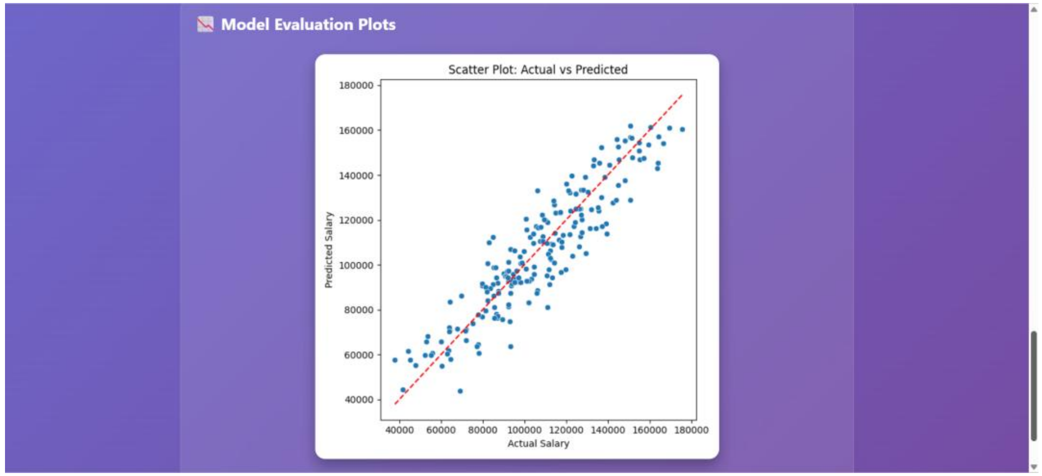
RESULT



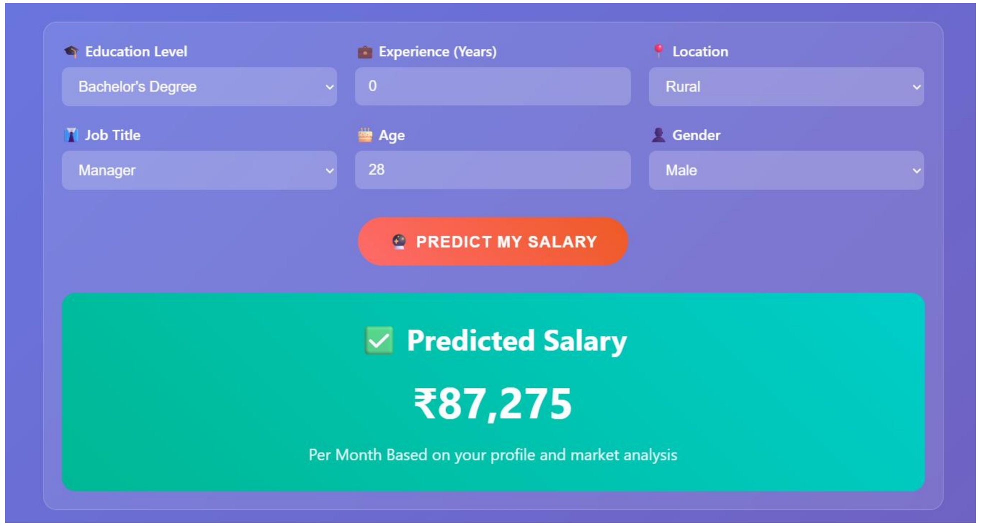
Model Evaluation Plots



Model Evaluation Plots



Model Evaluation Plots



Education Level: Bachelor's Degree

Experience (Years): 0

Location: Rural

Job Title: Manager

Age: 28

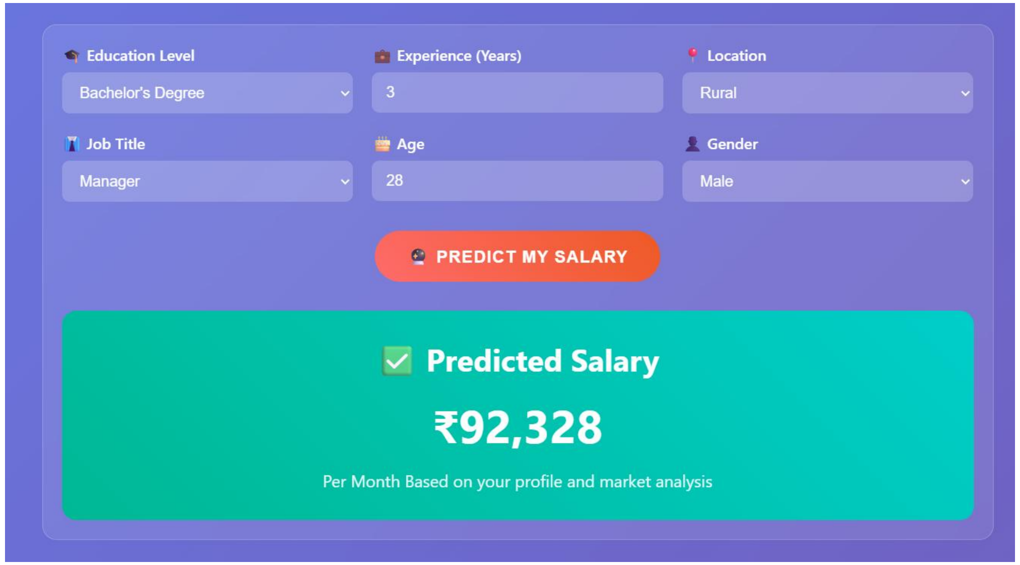
Gender: Male

✅ Predicted Salary

₹87,275

Per Month Based on your profile and market analysis

RESULT



Education Level: Bachelor's Degree

Experience (Years): 3

Location: Rural

Job Title: Manager

Age: 28

Gender: Male

✅ Predicted Salary

₹92,328

Per Month Based on your profile and market analysis

Github :- https://github.com/GourabGorai/IBMInternshipProject

CONCLUSION

The Salary Prediction Web Application successfully demonstrates the integration of machine learning with a user-friendly web interface to predict employee salaries based on various demographic and professional attributes. The Random Forest Regression model, combined with appropriate preprocessing and validation, delivers reasonably accurate predictions, as reflected by the R^2 score and error metrics. The inclusion of visual plots like actual vs. predicted salaries, residual distributions, and feature importance enhances interpretability for end-users.

- During the development, some key **challenges** were:
  - Ensuring logical consistency between age and experience inputs.
  - Managing invalid or missing entries in user-submitted data.
  - Generating meaningful insights from limited or noisy datasets.
  - Handling errors and edge cases during real-time predictions and plotting.

Despite these, the project achieved its core objective: providing reliable salary predictions with visual feedback and error-handling mechanisms.

REFERENCES

- Flask Documentation – https://flask.palletsprojects.com/
- Scikit-learn Documentation – https://scikit-learn.org/stable/
- Matplotlib Documentation – https://matplotlib.org/stable/contents.html
- Seaborn Documentation – https://seaborn.pydata.org/
- Stack Overflow – Community discussions and solutions related to Flask and ML deployment.
- Kaggle Datasets – For accessing various salary prediction datasets: https://www.kaggle.com/datasets
- W3Schools – For basic HTML and CSS used in Flask templates: https://www.w3schools.com/
- Python Official Documentation – https://docs.python.org/3/

THANK YOU