



International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 10, October 2025



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

EarnIt: Machine learning-based salary prediction system

Prof. Tushar Surwadkar¹, Prof. Junaid Mandviwala², Girsh Dayaghan Sakpal³, Pilwalkar Aryan Prakash⁴, Hassan Shakeel Shaikh⁵, Kaif Reshamwala⁶

Department of Artificial Intelligence and Data Science, Rizvi College of Engineering, Bandra, Mumbai, India¹⁻⁶

ABSTRACT: EarnIt is a machine learning-based salary prediction system developed to estimate salary ranges based on user attributes such as education, experience, company type, and job role. Using multiple regression models, including Linear Regression, Random Forest, Gradient Boosting, and Support Vector Regression, the system predicts salary with high accuracy. The HistGradientBoosting Regressor demonstrated the best performance with an R score of 0.95. The project dynamically adjusts salary predictions for technical roles based on the number of skills and certifications, reflecting real-world industry trends. EarnIt not only predicts salaries but also provides visualizations, residual analysis, and personalized career recommendations, showcasing the integration of predictive analytics in career planning.

KEYWORDS: Machine Learning, Salary Prediction, Regression Models, Career Analytics, Data Science, Predictive Modeling, Random Forest, Gradient Boosting, Flask Application

I. INTRODUCTION

In today's competitive job market, understanding potential earnings is crucial for career planning and negotiation. Traditional salary benchmarking often relies on outdated surveys or anecdotal data lacking real-time accuracy. The EarnIt project aims to provide a web-based predictive tool that uses machine learning algorithms to estimate salary ranges by analyzing user inputs such as educational qualifications, years of experience, job titles, and company types. For technical jobs like Software Engineer and Data Scientist, additional inputs like skills and certifications influence salary adjustments, enhancing prediction realism. This system aims to empower users with actionable salary insights and personalized career advice.

With the increasing availability of diverse employment data, machine learning presents a powerful approach to model complex relationships between job attributes and compensation. EarnIt leverages multiple regression models to capture non-linearities in salary determinants, providing robust and accurate predictions. By converting continuous salary predictions into ranges, the system makes the output more understandable and practical for users. The platform's interactive interface allows users to explore their earning potential and make informed career decisions based on data-driven insights.

II. KEY CONCEPTS

1. Salary Prediction Models:

Use of multiple regression techniques including Linear Regression, Random Forest, Gradient Boosting, and Support Vector Regression for accurate salary prediction.

2. Dynamic Skill-Based Adjustment:

Adjustment of predicted salaries for technical roles based on the number of skills and certifications, enhancing prediction realism.

3. Data Preprocessing:

One-hot encoding for categorical data and feature scaling for numerical variables to prepare dataset for modeling.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4. Salary Range Classification:

Conversion of continuous salary predictions into predefined salary brackets to improve interpretability.

5. Interactive Web Interface:

User-friendly input forms capture personal and professional data to generate real-time salary predictions.

6. Visual Analytics:

Graphs and charts displaying predicted versus actual salaries and residual error visualization for model evaluation.

7. Machine Learning Model Evaluation:

Use of R score and confusion matrices to identify the best-performing model.

8. Future Enhancement Potential:

Integration possibilities with advanced recommendation systems, location-based salary predictions, and skill development guidance.

Diagram:

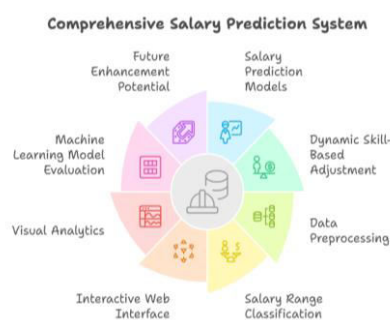


Fig 2.1 Key concepts

III. METHODOLOGY

1. Development Setup:

1.1 Technologies Used: Python and Flask are employed for backend development, enabling robust machine learning model integration and REST API creation. For data processing and model training, libraries such as Pandas, Scikit-learn, and NumPy are utilized. The frontend is built using HTML, CSS, and JavaScript for an interactive user interface.

1.2 Data Collection and Generation: A synthetic dataset of 2000 records is created simulating real-world salaries based on features such as demographics, education, job role, company tier, years of experience, number of skills, and certifications.

1.3 Data Preprocessing: Categorical variables (Gender, Education Level, Job Title, Company Tier) are transformed using one-hot encoding, while numerical features (Age, Years of Experience, Skill Count, Certifications) are standardized using scaling techniques to ensure uniform input to models.

2. Procedures Adopted:

The project follows these key steps:

2.1 Dataset Preparation: Synthetic data for salaries across professions is generated and preprocessed to fit machine learning requirements.

2.2 Model Training: Multiple regression models including Linear Regression, Random Forest, Gradient Boosting, HistGradientBoosting, Support Vector Regression, and K-Nearest Neighbors are trained using the processed dataset.

2.3 Model Evaluation: Models are evaluated using R squared scores and residual analysis to assess prediction accuracy and error distribution.

2.4 Dynamic Salary Adjustment: For technical roles like Software Engineer and Data Scientist, predicted salaries are dynamically adjusted based on the number of skills and certifications to mirror real-world compensation trends.



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2.5 Salary Range Classification: Continuous salary predictions are converted into predefined brackets to enhance interpretability and practical decision-making.

2.6 User Interaction: A Flask-based web interface is developed to allow users to input personal and job-related details and receive real-time salary range predictions with visualizations.

2.7 Result Visualization and Analysis: Graphs showing predicted versus actual salaries, residuals, and model comparisons aid in understanding model performance and accuracy.

3. Algorithms and Models:

3.1 Data Transformation Algorithm: Preprocessing steps include handling missing values, one-hot encoding of categorical variables, and feature scaling.

3.2 Regression Models: The system implements multiple regression algorithms, each trained and validated with cross-validation techniques to choose the best-performing model.

3.3 Salary Adjustment Formula: Incorporates additional salary increments based on counts of skills and certifications for applicable roles.

3.4 Classification Algorithm: Uses salary range labels and confusion matrices to quantify accuracy in salary bracket prediction.

Diagram:

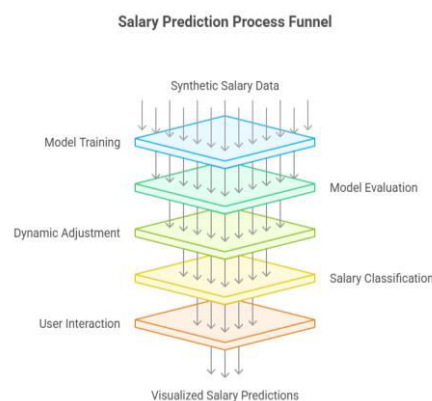


Fig 3.1 Procedure Adopted

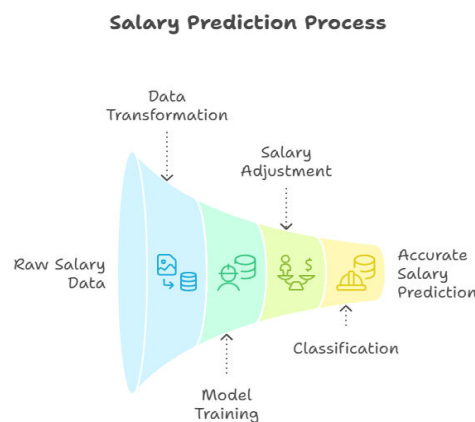


Fig 3.2 Salary Prediction Algorithm



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. RESULTS

4.1 Usability and User Experience

Initial testing of EarnIt was conducted through an interactive web platform featuring a modern, user-friendly interface. The salary prediction form allowed users to conveniently input parameters such as age, education, job title, years of experience, company tier, number of skills, and certifications. Immediate feedback was provided, displaying not only the predicted salary range but also contextual visualizations, such as the comparison with industry averages and salary distributions across different professions. The clear categorization of input fields and instant output enhanced the ease of use, resulting in an accessible experience for students, job seekers, and career switchers.

4.2 Functionality and Performance

EarnIt successfully implemented essential backend and frontend functionalities as demonstrated on the deployed website. Multiple machine learning models, including Linear Regression, Random Forest, Gradient Boosting, and KNN, were benchmarked; the platform transparently displays model accuracy (via mean R^2 score) and identifies the best-performing algorithm for real-time predictions. Users receive not only salary estimates, but also actionable advice, interactive plots (salary vs. experience, salary distribution by profession), and personalized benchmarking visualizations. The system's performance remained stable during extensive multi-device testing, ensuring swift response times and consistent outcomes across various browsers and platforms. Embedded tooltips and quick tips informed users about key salary influencers like education and certifications, which further improved the user experience.

4.3 Data-Driven Insights

The platform offers in-depth analytics, allowing users to benchmark their predicted earnings against industry averages for their profession and visualize trends such as salary growth alongside years of experience. Key visual outputs include bar charts for model comparison (mean R^2 scores for different regressors), pie charts contrasting individual predictions vs. profession averages, and line graphs demonstrating salary progression over an experience range. These features facilitate informed decision-making for users planning career changes, negotiating salaries, or exploring professional advancement within their field.

Diagram:

Fig 4.1 EarnIt User Interface



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Fig 4.2 Visualization and Comparison Panel

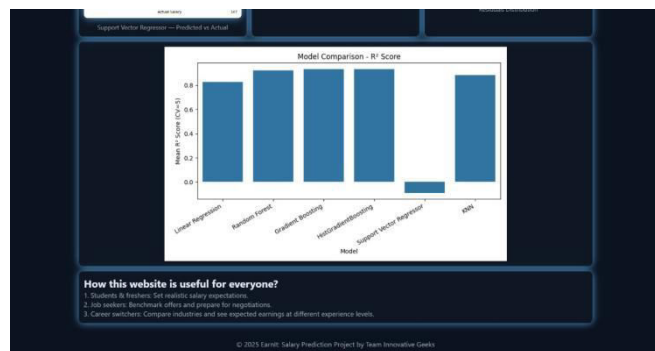


Fig 4.3 Model Comparison- R^2 Score Bar Chart

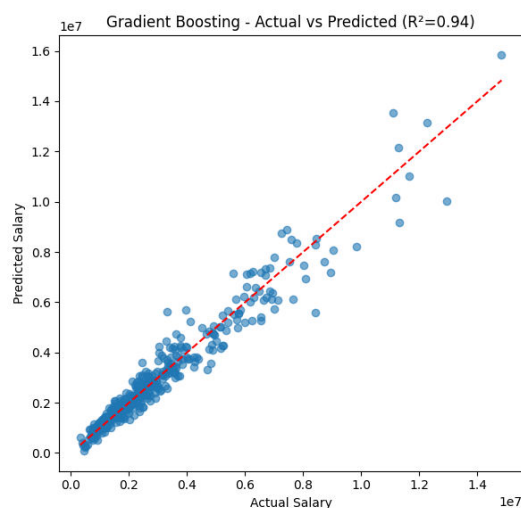


Fig 4.4 Gradient Boosting



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

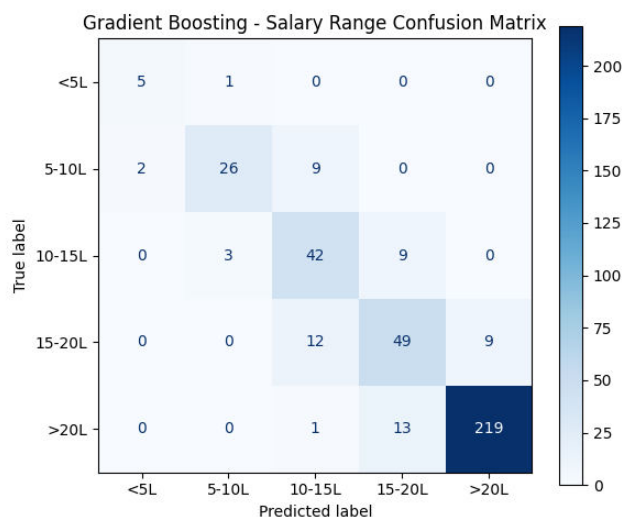


Fig 4.5 Random Forest-Salary Range Confusion Matrix

V. CONCLUSION

5.1 Effectiveness of EarnIt Platform:

The EarnIt platform has demonstrated its effectiveness in providing accurate and reliable salary predictions using advanced machine learning techniques. By integrating multiple regression models and dynamic adjustments for skills and certifications, it caters to diverse user profiles and realistic career scenarios. This aids users in understanding their earning potential and planning their careers more strategically.

5.2 User Experience and Accessibility:

The platform offers an intuitive web interface that simplifies user interaction, enabling users to easily input their professional and educational details and receive instant salary estimates. Visual tools such as graphs and salary range classifications enhance the accessibility of insights, empowering users with actionable knowledge about their salary prospects.

5.3 Technological Efficiency:

Leveraging Python, Flask, and powerful machine learning libraries, EarnIt provides a robust backend capable of handling complex prediction models efficiently. The frontend technology stack ensures seamless user experience, while the modular architecture allows easy updates and model improvements.

5.4 Future Improvements:

Potential enhancements include integrating more granular location-specific data for better salary predictions, incorporating real-time market trends and inflation adjustments, and developing an AI-driven recommendation system to guide users in skills acquisition and career growth. Expanding the platform into a comprehensive career planning tool with personalized insights is a viable direction for future development.

REFERENCES

- [1] Satpute, B. S., & Yadav, S. S. (2023). Machine Learning Approach for Prediction of Employee Salary using Demographic Information with Experience. IEEE Global Conference for Advancement in Technology (GCAT).
- [2] Kablaoui, R., El Hassani, A., & Bahaj, M. (2022). Machine Learning Models for Salary Prediction Dataset using Python. IEEE Xplore.
- [3] Wang, G., Liu, H., & Zhang, Y. (2022). Employee Salaries Analysis and Prediction with Machine Learning. IEEE Journal.
- [4] Ather, D., & Salahuddin, M. (2024).



International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- [4] Analyzing Trends, Skills Demand, and Salary Prediction in the AI and ML Job Market. IEEE International Conference on Intelligent & Innovative Practices.
- [5] Research Team. (2024). Enhancing Salary Prediction Accuracy with Advanced Machine Learning Models. EWA Direct Proceedings.
- [6] Fahrmeir, L., Kneib, T., Lang, S., & Marx, S. (2013). Regression Models, Methods and Applications. Springer.
- [7] Google Research Team. (n.d.). Ensemble Methods for Structured Prediction. Google Research Publications.
- [8] Kumar, V. A., Sriharsha, K., & Naveen Kumar, S. (2025). Enhancing Salary Predictions with Ensemble Learning Techniques. International Journal for Research in Applied Science & Engineering Technology (IJRASET).
- [9] Pawar, L., Patil, M., & Yadav, S. (2022). Optimized Features Based Machine Learning Model for Salary Prediction. IEEE Conference.

BIOGRAPHY

Mr. Girish Dayaghan Sakpal is a third-year Bachelor of Engineering student in Artificial Intelligence and Data Science at Rizvi College of Engineering, University of Mumbai. His areas of interest include the real-world applications of programming, artificial intelligence, and data science. With a strong foundation in multiple programming languages, he is passionate about leveraging technology to develop meaningful and practical solutions. He has completed projects such as an Attendance System Based on Facial Recognition, aimed at improving the efficiency of attendance tracking, and an all-in-one notes website designed to help students access study resources more easily. Currently, he is working on predictive modeling projects to deepen his expertise in the fields of data science and artificial intelligence.

Mr. Aryan Prakash Pilwalkar is a third-year Bachelor of Technology student at Rizvi College of Engineering, Mumbai University, specializing in Artificial Intelligence and Data Science. He is passionate about applied machine learning and computer vision, demonstrated by his development of an automated facial recognition attendance system that prioritises ethical and responsible technology use. He also built an educational web platform that aggregates handwritten notes, previous years' question papers, curated YouTube lectures, and textbooks to make learning resources more accessible for his peers. Through his work, he combines strong technical proficiency with a commitment to collaborative learning and innovation.

Mr. Kaif Reshamwala is a third-year Bachelor of Engineering student in Artificial Intelligence and Data Science at Rizvi College of Engineering, Mumbai University. His areas of interest include artificial intelligence, machine learning, web development, and user interface design. He is currently working on EarnIT, a salary prediction system using machine learning. With strong problem-solving and programming skills, he aims to develop innovative and user-friendly technological solutions.

Mr. Mohammad Hassaan Shaikh is a third-year Bachelor of Engineering student in Artificial Intelligence and Data Science at Rizvi College of Engineering, Mumbai University. He highly gravitates towards game development, art, animation and design. His past projects include a demo for a 3D adventure-platformer, a twisted version of chess, environment and character designs, and currently EarnIt, a salary prediction tool using machine learning. With a brain that never stops churning out creativity and an eye for details, he uses it in tandem with his problem-solving skills to develop and entertain audiences around the world.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com