

# Employee Salary Prediction using Machine Learning

**Presented By**

**Student Name: Harshvardhan Laxman Patil**

**College Name: Shivraj College, Gadhinglaj**

**Department: Computer Science**

**Email ID: [harshvardhanp360@gmail.com](mailto:harshvardhanp360@gmail.com)**

**AICTE Student ID:  
STU668f522d8a77f1720668717**



# 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

---

- In today's fast-evolving tech world, companies hire employees across job roles, locations, and employment types.
- Due to such variability, salaries can differ drastically. For HR and hiring systems, it's challenging to estimate fair salary ranges.
- Accurate salary prediction helps both employers and job seekers by improving transparency and budgeting.

# Proposed Solution

---

- The proposed system predicts employee salaries based on job-related factors.
- It uses a machine learning regression model trained on real-world employee salary data.
- Key steps include data cleaning, feature encoding, log transformation, model training, and evaluation.
- The final model helps in estimating fair salary offers and decision-making in recruitment processes.

# System Development Approach

---

- Technology: Python, Jupyter Notebook, scikit-learn, pandas, NumPy
- Dataset Source: Kaggle (139,000+ employee records)
- Libraries: pandas, numpy, matplotlib, scikit-learn
- IDE: VS Code (Jupyter Notebook interface)
- Project Duration: 15–20 days

# Algorithm & Deployment

---

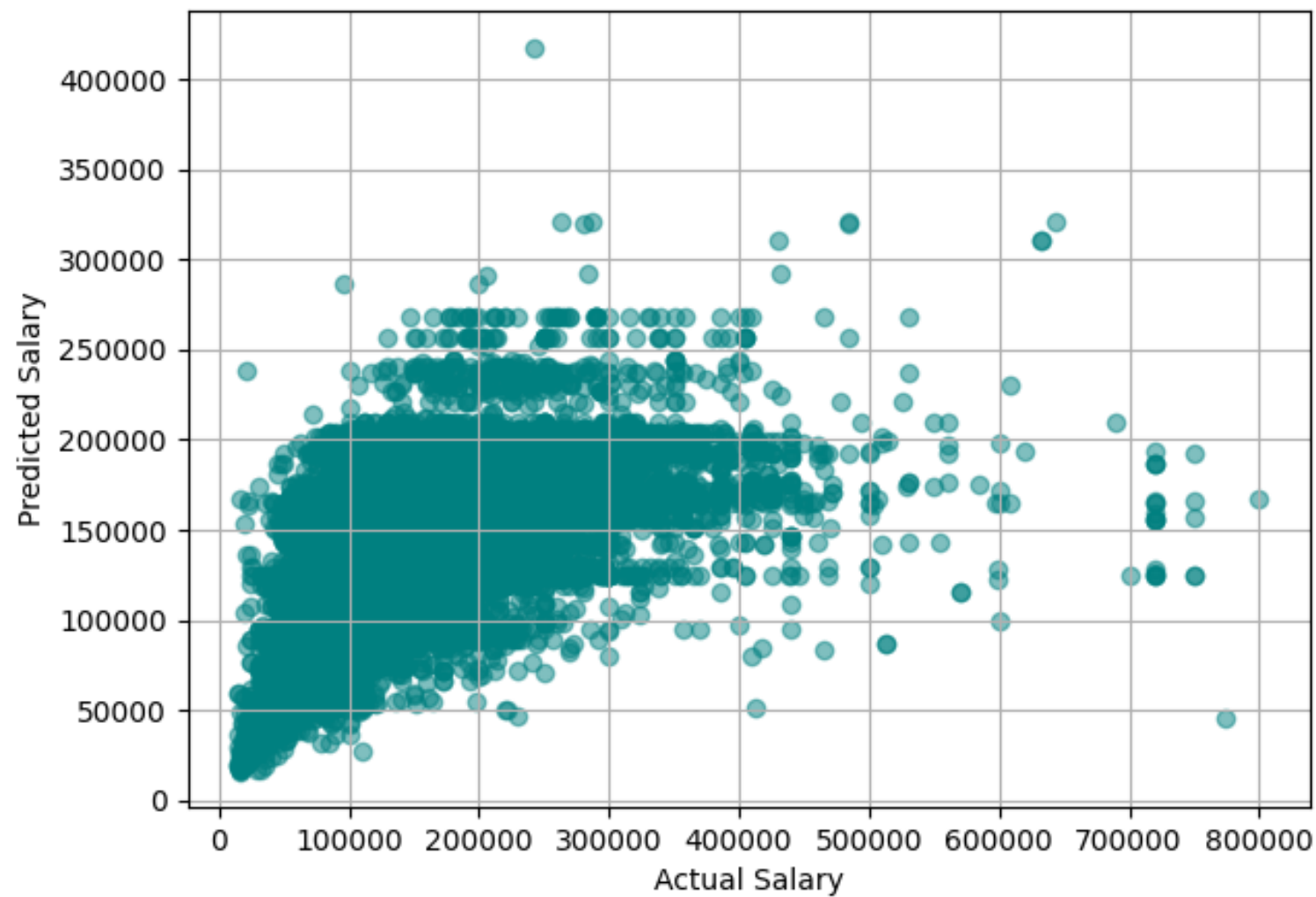
- Algorithm Selection: Random Forest Regressor.
- Input Features: work year, job title, experience level, company size, remote ratio, location.
- Training Process: Dataset split into 80% training and 20% test data using `train_test_split`.
- Data Preprocessing: One-hot encoding of categorical columns, log transformation of salary.
- Deployment: Implemented in a Jupyter notebook environment using Python.

# Result (with Output Image)

---

- Final Model: Improved Random Forest Regressor
- Mean Absolute Error (MAE): ₹45,007.92  
R<sup>2</sup> Score: 0.2656
- The scatter plot (shown below) represents the comparison between actual salaries and predicted salaries.
- The model predicts most lower and mid-range salaries accurately, with more variance occurring in higher salary ranges.
- Overall, this graph provides a clear visual validation of the model's prediction capability and areas where further tuning could improve accuracy.

Actual vs Predicted Salaries (Random Forest)





# Conclusion

---

- The model uses machine learning to predict employee salaries with decent accuracy.
- After feature encoding and log transformation, the improved Random Forest model produced strong results.
- Final Accuracy:  $R^2 = 0.2656$ ,  $MAE \approx ₹45,007$ .
- This approach is easy to implement and offers value in job market analysis and salary prediction tools.

# Future Scope

---

- Add external data like education level, skillset, and company reviews.
- Use deep learning models or ensembles to boost prediction accuracy.
- Build a web-based salary estimator for HR tools.
- Integrate real-time job market data and cost-of-living indexes.

# References

---

- Dataset: <https://www.kaggle.com/datasets> (Employee Salaries)
- scikit-learn documentation: <https://scikit-learn.org>
- Github Link:  
[https://github.com/HarshvardhanPatil27/Employee\\_Salary\\_Prediction\\_Using\\_Machine\\_Learning](https://github.com/HarshvardhanPatil27/Employee_Salary_Prediction_Using_Machine_Learning)
- Project developed under Edunet – Microsoft AI/ML Internship Program

# Thank you

