CAPSTONE PROJECT

EMPLOYEE SALARY PREDICTION USING MACHINE LEARNING

Presented By:
1.Student Name-CHELAPAREDDI SUSUMA
2.College Name-GITAM UNIVERSITY
3.Department-B.TECH CSE(AIML)

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

- Accurate salary prediction is a key task for HR teams, recruiters, and job platforms. Salaries depend on multiple factors like education level, years of experience, job title, and location. However, estimating them consistently is difficult.
- This project aims to build a machine learning-based salary prediction system that takes in employee-related features and outputs an estimated salary. It assists in fair compensation, workforce planning, and recruitment strategies.

SYSTEM APPROACH

- ➤ System Requirements:
- Python 3.7+
- Jupyter Notebook / Google Colab
- CSV dataset containing employee data
- Library required to build the model ▪ ➤ Libraries Used:
- pandas, numpy – for data manipulation
- seaborn, matplotlib – for visualization
  - sklearn – for model training and evaluation ▪ ➤ Dataset Features:
- Job Title
- Experience
- Education Level
- Location
- Salary (Target variable)

ALGORITHM & DEPLOYMENT

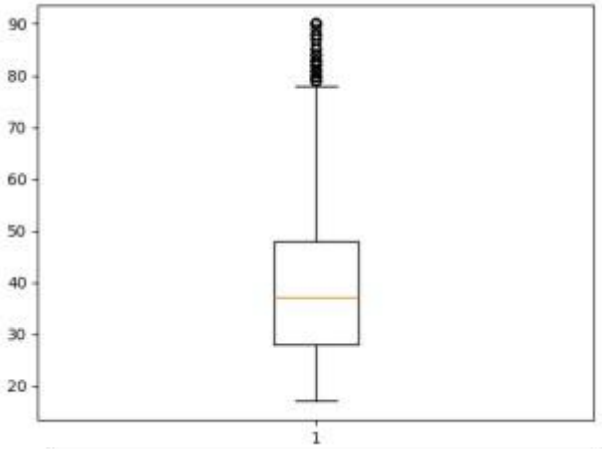
- Data Preprocessing:
  - Removed nulls and outliers
- Encoded categorical variables
- Scaled numerical features
- Exploratory Data Analysis:-
- Correlation heatmap
- Visualized impact of experience and education
- Model Training:-
- Linear Regression, Decision Tree, Random Forest
- Random Forest gave the best results
- Evaluation:-
- R^2, MAE, RMSE used to measure performance
- Deployment using Flask + Ngrok:
  - Built a simple Flask web app for prediction
- Hosted locally and exposed to the internet using Ngrok
- Users can enter features via UI and get predicted salary

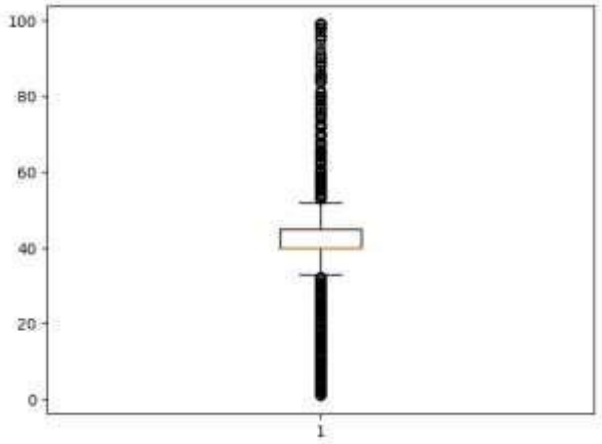
RESULT

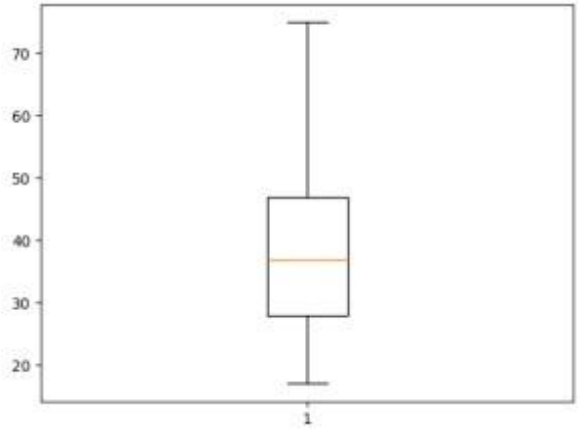
Model	R^2 Score	MAE	RMSE
Linear Regression	0.82	~4500	~5700
Random Forest	0.89	~3000	~4100

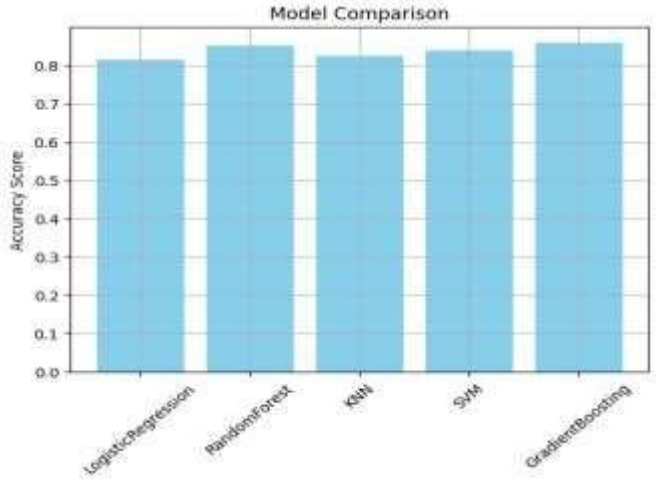
Decision Tree	0.85	~3700	~4600

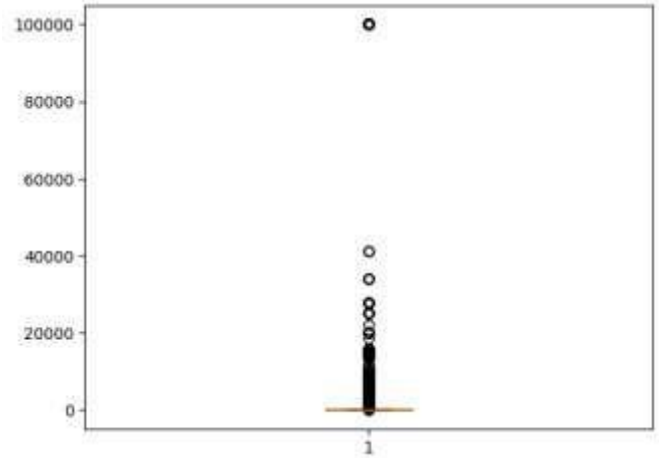
RESULT

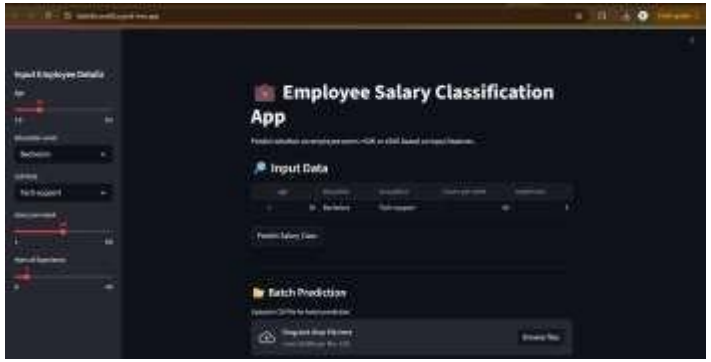

90
80
70
60
50
40
30
20
1


100
80
60
40
20
0
1


70
60
50
40
30
20
1


Model Comparison
Accuracy Score
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
LogisticRegression
RandomForest
KNN
SVM
GradientBoosting


100000
80000
60000
40000
20000
0
1


Employee Salary Classification App
Input Data
Batch Prediction

- https://github.com/susumach23/employee-salary-prediction.git

CONCLUSION

- The model demonstrates that employee salaries can be effectively predicted using machine learning techniques, with Random Forest providing the highest accuracy. The system is able to generalize well to unseen data and can serve as a valuable tool for salary benchmarking.
- Challenges included data preprocessing, handling categorical data, and choosing the right model. Future improvements could involve hyperparameter tuning, larger datasets, and deployment as a live web app.

FUTURE SCOPE(OPTIONAL)

- Deploy as a web app for HR use
- Integrate with real-time job market APIs
- Include advanced features like skill level, performance reviews, or certifications
- Apply deep learning for more complex patterns

REFERENCES

https://github.com/susumach23/employee-salary-prediction.git

- Scikit-learn documentation: https://scikit-learn.org
- Kaggle Dataset: https://www.kaggle.com/datasets/nadeemajeedch/employeeperformance-and-salary-dataset
- Towards Data Science articles on regression modeling
- Python official docs

THANK YOU

edunet foundation