

Salary Prediction System using Machine Learning

Ms. Gurpreet¹, Kunal Bistania², Sarthak Mahendroo³, Neha Dogra⁴, Paridhi Tuteja⁵

1 Assistant Professor, Dept. of Computer Science and Engineering, HMR Institute of Technology & Management, New Delhi, India

2345 Students, B. Tech, Dept. of Computer Science and Engineering, HMR Institute of Technology & Management, New Delhi, India

1. ABSTRACT

Salary prediction is a popular problem among the Data Science community for complete beginners. Through this regression analysis, we found a perfect linear relationship between the salary and the job experience of the people. It means more job experience results in a higher salary. This research paper presents a software Salary Prediction System that aims to address the challenge of accurately estimating salaries in the software industry. The system provides valuable insights for employers and employees in the face of increasing demand for skilled software professionals and the need for fair compensation.

2. OBJECTIVE

The objective of the Salary Prediction System is to create an accurate and reliable model that can estimate the salaries of software developers objectively. This system aims to provide valuable insights based on

data analysis, benefiting both job seekers and employers in the software development industry. The key objectives of the system are as follows:

1. **Accurate Salary Estimation:** Develop a predictive model capable of accurately estimating the salaries of software developers. The model will consider a range of factors, including experience, education level, location to generate precise salary predictions.

2. **Data-Driven Insights:** Provide job seekers with valuable insights into the expected salary range based on their qualifications, experience, and geographical location. This information will empower individuals to make well-informed decisions regarding job offers, negotiate salaries effectively, and plan their career growth.

3. **Employer Salary Benchmarking:** Assist employers in benchmarking their salary offerings against industry standards. By

analyzing market trends and comparing salaries for similar positions, employers can ensure competitive compensation packages that attract and retain talented software developers.

3. METHODOLOGY

1. Data Collection: Gather a comprehensive dataset consisting of relevant information such as software developer profiles, their qualifications, experience, skills, educational background, location, and salary data. This dataset should also include additional factors that may influence salaries, such as company size, industry, and market trends.

2. Data Preprocessing: Cleanse and preprocess the collected data to ensure accuracy and consistency. This step may involve removing duplicates, handling missing values, standardizing data formats, and transforming categorical variables into numerical representations.

3. Feature Selection and Engineering: Identify the most relevant features that contribute significantly to salary predictions. Use domain knowledge and statistical analysis techniques to select the appropriate set of features. Additionally, engineer new features that capture valuable information, such as combining experience and

educational qualifications to create a composite metric.

4. Model Selection: Explore different machine learning algorithms suitable for regression tasks. Compare and evaluate their performance using techniques such as cross-validation and metrics like mean squared error (MSE) or root mean squared error (RMSE). Consider algorithms such as linear regression, decision trees, random forests, support vector regression, or gradient boosting.

5. Model Training: Split the dataset into training and validation sets. Train the selected model(s) on the training set and optimize the hyperparameters to achieve the best performance. Validate the model(s) using the validation set to ensure generalizability and prevent overfitting.

6. Model Evaluation: Assess the performance of the trained model(s) using appropriate evaluation metrics such as mean absolute error (MAE), R-squared value, or mean percentage error (MPE). Compare the performance against industry benchmarks and baseline models to determine the effectiveness of the developed salary prediction system.

7. Deployment: Implement the trained model(s) into a user-friendly application or

system interface that allows users to input their parameters and receive personalized salary predictions. Ensure the system is scalable, secure, and capable of handling real-time data inputs.

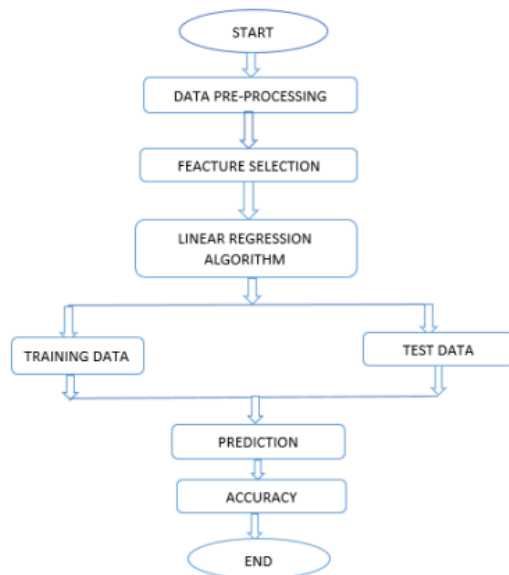


Fig1. Architecture Diagram

4. INTRODUCTION

In today's dynamic landscape of software development, having a comprehensive understanding of salary trends and accurate salary predictions is paramount for both job seekers and employers. With the increasing demand for skilled software developers, it has become essential to provide objective and data-driven insights into compensation levels. To fulfill this need, the Salary Prediction System has emerged as a valuable

tool.

The Salary Prediction System aims to provide individuals and organizations with reliable and accurate estimations of software developer salaries. By considering a wide range of factors, including experience, skills, education level, location, this system utilizes advanced predictive modeling techniques to generate valuable insights into expected salary ranges.

However, the amount of salary depends on the reputation of the school's name, the student's academic background, and other activities while at the university. Therefore, to estimate the salary, the academic record is used as a reference for the salary. The model is used as a tool to show the approximate salary by taking various factors into considerations.

5. EXPERIMENTATION

1. Linear Regression:

- Linear regression is a statistical modelling technique used to establish a relationship between a dependent variable (target variable) and one or more independent variables (features) by fitting a linear equation.
- The Linear Regression class from the sklearn library can be used to estimate the coefficients of the linear equation

and predict the salary based on years of experience.

2. Pandas:

- Pandas is a powerful library for data manipulation and analysis. It provides data structures and functions to handle structured data, such as CSV files, in a tabular format (DataFrame).
- In the salary prediction system, Pandas can be utilized to load the dataset, separate features and the target variable, and perform various data pre-processing tasks.

3. NumPy:

- NumPy is a fundamental library for numerical computing in Python. It supports large, multi-dimensional arrays and matrices, along with a collection of mathematical functions for efficient computations.
- NumPy is commonly used for mathematical operations, array manipulation, and numerical computations involved in machine learning algorithms.

4. Matplotlib:

- Matplotlib is a widely used plotting library in Python. It provides a flexible set of functions for creating various types of plots and

visualizations.

- In the salary prediction system, Matplotlib can be employed to visualize the training set results by creating a scatter plot of the actual data points and plotting the regression line.

5. Pickle:

- Pickle is a Python module used for object serialization, allowing the saving and loading of Python objects (e.g., trained models) to/from disk.

6. Scikit-learn (sklearn):

- scikit-learn is a widely used machine learning library in Python. It offers a diverse set of tools and algorithms for tasks such as classification, regression, clustering, etc.
- In the salary prediction system, the Linear Regression class and the `train_test_split` function from sklearn are utilized. Linear Regression is used to train the model, while `train_test_split` helps split the dataset into training and testing sets.

These components work together to build a salary prediction system that employs linear regression to estimate salaries based on years of experience. Pandas and NumPy facilitate data pre-processing, Matplotlib enables

visualization, and pickle allows for saving and loading the trained model.

6. RESULTS AND DISCUSSION

Representative image can be seen from this application and can also estimate a point from the location and calculate the salary. Therefore, salary surveys are categorized according to their sources of information - getting information from companies or collecting information from employees. Auditors are awarded for efforts to obtain the most important information by all means possible.

First we have the linear function. "LinearRegression()" in this function creates a straight line between points.

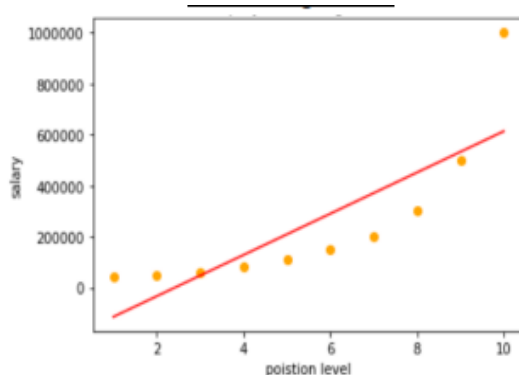


Fig2. Linear Regression

In Figure 2, we can see that there is no reference point. That's why we now use "polynomial properties" for curves. Here "degree = 6" refers to the smoothness of the curve as above, for curves

"lg2.predict(poly_reg.fit_transform(x))" is used instead of straight lines.

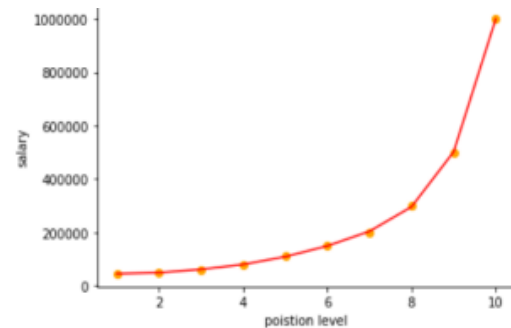


Fig 3: Polynomial Regression

7. CONCLUSIONS

The implementation of the Salary Prediction System for software developers represents a significant advancement in the field of employment analysis and planning. By leveraging the capabilities of data analysis and predictive modelling, this system provides accurate and unbiased salary estimations, benefiting both job seekers and employers alike.

For job seekers, the system offers valuable insights into the anticipated salary range based on their qualifications, experience, and geographical location. Armed with this information, individuals can make more informed decisions when evaluating job offers, effectively negotiate salaries, and strategically plan their career progression.

8. REFERENCES

1. Krishna Gopal, Ashish Singh, Harsh Kumar, Dr. Shrddha Sagar, "Salary Prediction Using Machine Learning", June 2021| IJIRT | Volume 8 Issue 1.
2. Sayan Das, Rupashri Barik, Ayush Mukherjee, "SALARY PREDICTION USING REGRESSION TECHNIQUES", SSRN Electronic Journal, January 2020.
3. Tiasa Mukherjee, MS. B. Satyasaivani, "Employee's Salary Prediction" International Journal of Advance Research, Ideas and Innovations in Technology | Volume 8, Issue 3.
4. Susmita Ray, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (ComIT-Con), India, 14th -16th Feb 2019.
5. A.L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection", IEEE Communication Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176.
6. Tzanis, George, et al. "Modern Applications of Machine Learning." Proceedings of the 1st Annual SEERC Doctoral Student Conference–DSC. 2006.
7. Mitchell, Tom Michael. The discipline of machine learning. Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2006.
8. Pornthep Khongchai, Pokpong Songmuang, "Improving Students' Motivation to Study using Salary Prediction System" 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE).
9. Navyashree M, Navyashree M K, Neetu M, Pooja G R, Arun Biradar "Salary Prediction in It Job Market", International Journal of Computer Sciences and Engineering Vol.-7, May 2019.
10. Pornthep Khongchai, Pokpong Songmuang, "Improving Students Motivation to Study using Salary Prediction System" 2016 13th

International Joint Conference on Computer Science and Software Engineering (JCSSE).

11. Phuwadol Viroonluecha, Thongchai Kaewkiriya, "Salary Predictor System for Thailand Labour Workforce using Deep Learning" The 18th International Symposium on Communications and Information Technologies (ISCIT 2018).
12. A.W. Husain, N.A Rashid and A.M. Shahiri, A Review on Predicting Students Performance using Data Mining Techniques Procedia Computer Science 72:414-422, 2015.
13. Richard A Huebner, A Survey of Educational Data-Mining Research, Academic and Business Research Institute, 2013.
14. Pornthep Khongchai, Pokpong Songmuang, improving students' motivation to study using salary prediction system, 13th International Joint Conference on Computer Science and Software Engineering, 2016.
15. Karlar Hamlen and William A. Hamlen, Faculty Salary as a predictor of student outgoing salaries from MBA programs, Journal of Education for Business, 2016.
16. Rajveer Singh, A Regression Study of Salary Determinants in Indian Job Markets for Entry Level Engineering Graduates, Masters Dissertation. Dublin Institute of Technology, 2016.
17. C.C. Hung, E.-P. Lim, On Aggregating Salaries of Occupations from Job Post and Review Data", IEEE Access, Volume 9, 2021.
18. 21. Dr. Kamaljit I. Lakhtaria, Bhaskar Patel and S.G. Prajapati, Efficient classification of data using decision Tree, Bonfring international journal of data mining 2 (1), 06-12, 2012.
19. Phuwadol Viroonluecha, Thongchai Kaewkiriya, "Salary Predictor System for Thailand Labour Workforce using Deep Learning" The 18th International Symposium on Communications and Information Technologies (ISCIT 2018).