International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Salary Prediction using Linear Regression

Gauri Shinde, Swarup Shinde, Sujal Sawant, Janhvi Tambolkar

Department of Multidisciplinary Engineering, Vishwakarma Institute of Technology, Bibwewadi, Pune

Abstract: The goal of this project is to predict the salary of employees based on their years of experience using a simple linear regression model. The input data for the predictive model is taken as years of experience which is the basis. By using years of experience as the input data we get salary as the output. In the predictive model Linear Regression method is used as there is only one variable i.e. years of experience. With the help of this model by entering the years of experience we will be getting salary as the output. The dataset used in this study includes information on the years of experience andsalaries of employees in a particular organization. The data is preprocessed by checking for missing values, outliers, and transforming the variables as necessary. We then perform a simple linear regressionanalysis on the dataset to investigate the relationship between years of experience and salary. The resultsshow a strong positive linear relationship between the two variables, with a high coefficient of regression (R-squared) of 0.96, indicating that 96% of the variation in salary can be explained by the years of experience.

Keywords: Linear regression, Salary Prediction, Regression Coefficient (R^2)

Introduction

In today's world, a significant factor driving employee turnover is their salary. Employees often switch companies in pursuit of the desired compensation they expect. This trend not only causes losses for the company but also prompts the need for a solution. We have conceived an idea: What if employees could receive the salary they anticipate from their current organization? In this highly competitive environment, everyone has elevated expectations and goals. However, it is not feasible to provide arbitrary salary figures to all employees; rather, a system is required to assess an employee's suitability for their desired compensation. While an exact salary cannot be determined, it is possible to make predictions based on specific datasets. Predictions involve making assumptions about future events.

Machine Learning is an area of research that empowers computer systems to enhance their performance in a particular task without direct programming. This is accomplished by training the system using a dataset, enabling it to discern patterns and connections within the data. Subsequently, the system applies this acquired knowledge to make predictions or take actions in response to new inputs. Machine Learning encompasses diverse categories such as reinforcement learning, supervised learning, unsupervised learning and semi-supervised learning.

Predicting salaries through machine learning involves training a model using historical salary data to estimate future earnings for individuals. This process utilizes diverse supervised learning algorithms like linear regression, decision trees, or random forest. The model is trained on a dataset comprising various features such as job title, years of experience, education level, location, and other pertinent details, along with the target variable of salary. After training, the model becomes capable of making predictions on new data, such as forecasting the salary of a job candidate based on their experience and education level. The accuracy of the model relies on the quality of the training data and the chosen algorithm, necessitating careful validation and potential adjustments to enhance performance.

Literature Review

Krishna Gopal1 et al, [1] This paper aims at predicting the salary of the school students which they will be getting after completing their course. Because of this they will be knowing in which qualities they are lacking behind and they can improve those qualities which in case will contribute in their salary. By using data processing the model is build the results of school students and students who have graduated are compared .In this project they have used 10-fold cross-validation and data mining techniques.

Prof. D. M. Lothe1 et al, [2] In this paper a salary prediction model is build using linear regression algorithm. They have used 5 features like type of degree, subject specialization, and years of experience, type of industry to predict the salary of employee they have first performed on two variables i.e of one independent and one dependent years of experience and salary respectively with model accuracy of 96% to 98%.

Sayan Das et al, [3] The objective of this paper is to forecast an individual's salary at a specific future point in time. The visualization of salary prediction entails the creation of a computerized system that manages the progression of salary growth in various domains, enabling the anticipation of salaries over specific time periods. This application leverages the salary database of an organization to generate a graphical representation based on the available data. By examining the salary fields and importing the necessary graph, it facilitates visual observation. Moreover, employing a prediction algorithm, it enables the estimation of salaries for specific time intervals. Furthermore, this approach can be extended to encompass other impactful predictive applications as well.

U Bansal et al, [4] The work done in this paper is to predict the salaries of employee and real estate prices. The salaries are predicted on the basis of years of experience and many more factors. Real estate prices are predicted on the basis of land area, number of bedrooms, etc. Simple linear regression and Multiple linear regression techniques are used.

Methodology

The prediction models use the historical data to make predictions. The raw data of employee job experience and salaries is collected and it is used to train the model . When we will be giving the new input the model will predict our salary.

In this project various steps are involved the data science pipeline includes following steps :

1. Data collection
Unprocessed data which need to be processed Unformatted excel file, abinary file generated by a machine, Hand entered No's (Readings) you collected.

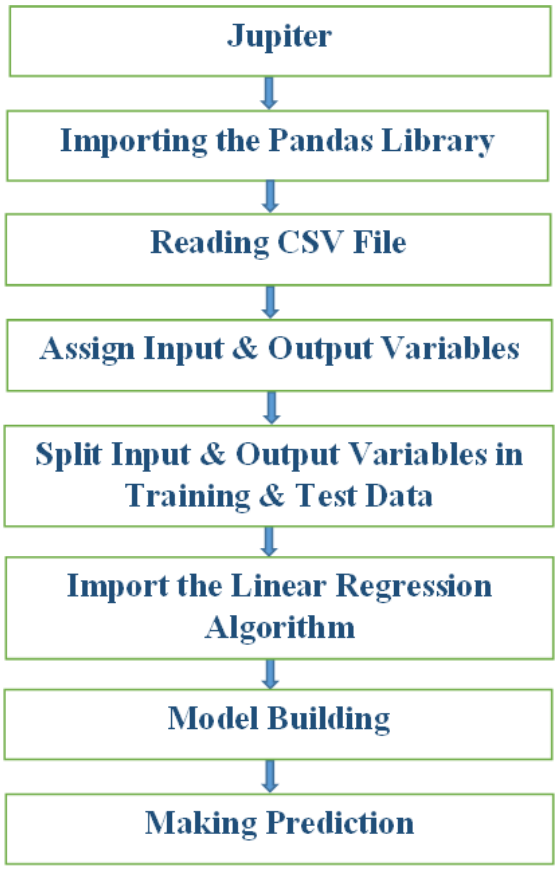
2. Data cleaning
Removal of unwanted information, handling missing data, fixing structural errors, managing unwanted outliers

3. Data visualization
Visual representation in the forms of graphs, charts, maps

4. Model Building
Various algorithms are used which best suits the model like linear regression etc

5. Model Deployment
Last stage of process where the model is ready and available for predictions

Detailed Steps



Results

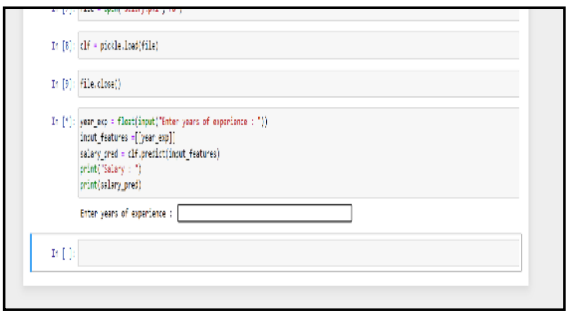


Fig.1.Enter the year experience

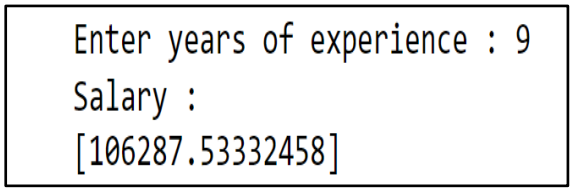
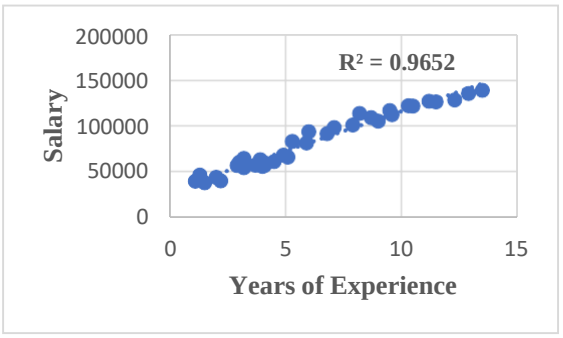
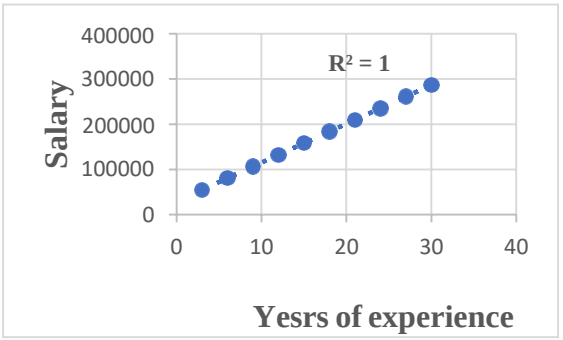


Fig.2.Get Salary as output result



Graph.1.Regression coefficient for input data



Graph.2.Regression coefficient for Predicted Data

Conclusion

Based on the analysis of the data, we found that there is a significant positive linear relationship between the years of experience and salary of employees. The linear regression model generated a high coefficient of determination (R-squared) of 1 for the predicted data, indicating that 100 % of the variation in the dependent variable (salary) canbe explained by the independent variable (yearsof experience). Therefore, we can conclude that years of experience is a strong predictor of salary for employees, and this relationship can be effectively modelled using a simple linear regression model. This information can be used to help organizations make informed decisions regarding employee compensation and career development.

Acknowledgement

We would like to thank our project guide Prof. Palomi Gawali for extending her guidance and support throughout the project.

References

1. *Krishna Gopal1 et al "Salary Prediction Using Machine Learning" © June 2021| IJIRT | Volume 8 Issue 1 | ISSN: 2349-6002.*
2. *Prof. D. M. Lothe1 et al "Salary Prediction Using Machine Learning" || Volume 6 || Issue 5 || May 2021 || ISSN (Online) 2456-0774 INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS IMPACT.*
3. *Sayan Das et al "SALARY PREDICTION USING REGRESSION TECHNIQUES"*
4. *Article in SSRN Electronic Journal · January 2020 DOI: 10.2139/ssrn.3526707.*
5. *U Bansal et al, "Empirical analysis of regression techniques by house price and salary prediction" To cite this article: U Bansal et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1022 012110.*