Salary Predictor System

Abstract:

In today's world, salary is the primary source of motivation for many regular employees, which makes salary prediction very important for both employers and employees. Higher Salaries motivate employees to apply for a company with more determination whereas low salaries discourage employees from spending more time with their application. It also helps the employer to understand what an employee is expecting a salary range from them. This is the major reason why we have created a model that quite accurately predicts the salary of a company, which will benefit both employers and employees immensely.

Methodology:

We used two different algorithms namely Logistic Regression and KNN algorithm. The main reason for using Logistic Regression was that it is one of the most efficient algorithms when the different outcomes or distinctions represented by the data are linearly separable. Also, we used the KNN algorithm as it provides very accurate predictions with such data. We also decided to showcase a comparison between these two algorithms to see which algorithm works better. The result was in favor of the KNN algorithm.

Working of the model:

Initially, we took a dataset that we divided 80 % for training purposes and 20% for testing purposes. We trained the data using Logistic Regression and the KNN algorithm. We imported various libraries such as numpy, pandas,matplotlib, sklearn, etc. The training data included various parameters such as max salary, min salary, median salary, posting period, posting listed, etc. Later we created a Correlation Matrix to relate different variables to each other. After that, we tried to make predictions on the test data set using logistic regression first. Next, we did the same steps but with the KNN algorithm. Finally, we used the matplotlib library to graphically demonstrate the difference in the accuracy of the prediction made by our model.

Overview:

Our model was able to predict the salaries of companies. Using logistic regression, we found the accuracy to be 60% which is a little less but on the other hand KNN algorithm gave us an accuracy of a whopping 77%. If larger datasets were available, I am sure we would be able to get even better results because these algorithms work or learn better with more data present. Despite this, our model could be used by anyone because of its simplicity and easy steps to implement.

Conclusion:

We can conclude that we successfully developed a model for Salary prediction using two different algorithms. This model could be used for predicting the salaries of Fortune 500 companies as well as small-scale startups. We successfully implemented the machine learning concepts of Logistic Regression and the KNN algorithm. The model that we trained was quite successful in predicting the data that we wanted. Our Hypothesis was successfully proved, which was developing a computerized system to maintain all the daily work of salary growth of any given company in any field and can predict salary after a certain time.