

# SALARY PREDICTION USING MACHINE LEARNING

Prof. D. M. Lothe<sup>1</sup>, Prakash Tiwari<sup>2</sup>, Nikhil Patil<sup>3</sup>, Sanjana Patil<sup>4</sup>, Vishwajeet Patil<sup>5</sup>

Professor, Information Technology, JSPM's Rajarshi Shahu College of Engineering, Pune, India<sup>1</sup> UG Student, Information Technology, JSPM's Rajarshi Shahu College of Engineering, Pune, India<sup>2345</sup> dml.rscoe@gmail.com<sup>1</sup>, prakashtiwari468@gmail.com<sup>2</sup>, np0129999@gmail.com<sup>3</sup>, patilsanjanam@gmail.com<sup>4</sup>, svishpatil@gmail.com

\_\_\_\_\_ \*\*\*\_\_\_\_\_

Abstract: - Machine learning is a technology which allows a software program to became more accurate at pretending more accurate results without being explicitly programmed and also ML algorithms uses historic data to predicts the new outputs. Because of this ML gets a distinguish attention. Now a day's prediction engine has become so popular that they are generating accurate and affordable predictions just like a human, and being using industry to solve many of the problems. Predicting justified salary for employee is always being a challenging job for an employer. In this paper and proposing a salary prediction model with suitable algorithm using key features required to predict the salary of employee.

Keywords: - Machine Learning, Linear regression, Model selection, Supervised Learning

\*\*\*\_\_\_\_\_

## I INTRODUCTION

Now days, Major reason an employee switches the company is the salary of the employee.

Employees keep switching the company to get the expected salary. And it leads to loss of the company and to overcome this loss we came with an idea what if the employee gets the desired/expected salary from the Company or Organization. In this Competitive world everyone has a higher expectation and goals. But we cannot randomly provide everyone their expected salary there should be a system which should measure the ability of the Employee for the Expected salary.

We cannot decide the exact salary but we can predict it by using certain data sets.

A prediction is an assumption about a future event.

In this paper the main aim is predicting salary and making a suitable user-friendly graph. So that an Employee can get the desired salary on the basis of his qualification and hard work. For developing this system, we are using a Linear regression algorithm of supervised learning in machine learning.

Supervised learning is basically a learning task of a learning function that maps an input to an output of given example. In supervised learning each example is pair having input parameter and the desired output value.

Linear regression algorithm in machine learning is a supervised learning technique to approximate the mapping function to get the best predictions. The main goal of regression is the construction of an efficient model to predict the dependent attribute from a bunch of attribute variables. A regression problem is when the output value is real or a continuous value like salary.

## **II LITERATURE REVIEW**

- 1) Susmita Ray," A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th -16th Feb 2019 a brief review of various machine learning algorithms which are most frequently used to solve classification, regression and clustering problems. The advantages, disadvantages of these algorithms have been discussed along with comparison of different algorithms (wherever possible) in terms of performance, learning rate etc. Along with that, examples of practical applications of these algorithms have been discussed.[1]
- 2) Sananda Dutta, Airiddha Halder, Kousik Dasgupta," Design of a novel Prediction Engine for predicting suitable salary for a job" 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN) - focused on the problem of predicting salary for job advertisements in which salary are not mentioned and also tried to help fresher to predict possible salary for different companies in different locations. The corner stone of this study is a dataset provided by ADZUNA. model is well capable to predict precise value.[2]
- 3) Pornthep Khongchai, Pokpong Songmuang, "Improving Students' Motivation to Study using Salary Prediction System" proposed prediction model using Decision tree technique with seven features. Moreover, the result of the system is not only a predicted salary, but also the 3-highest salary of the graduated students which share common attributes to the users. To test the system's efficiency, they set up an experiment by using 13,541 records of actual graduated student data. The total result in accuracy is 41.39%.[3]



# || Volume 6 || Issue 5 || May 2021 || ISSN (Online) 2456-0774 INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

4) Phuwadol Viroonluecha, Thongchai Kaewkiriya," Salary Predictor System for Thailand Labour Workforce using Deep Learning" - used Deep learning techniques to construct a model which predicts the monthly salary of job seekers in Thailand solving a regression problem which is a numerical outcome is effective. We used five-month personal profile data from wellknown job search website for the analysis. As a result, Deep learning model has strong performance whether accuracy or process time by RMSE 0.774 x 104 and only 17 seconds for runtime.[4]

## **III INTRODUCTION**

In order to gain useful insights into the job recruitment, we compare different strategies and machine learning models. The methodology different phases like: Data collection, Data cleaning, Manual feature engineering, Data set description, Automatic feature selection, Model selection, Model training and validation, Model comparison.

We are focusing to develop a system that will predict the salary based on different parameters used in company and abovementioned methodology phases. Some of the parameters we collected from company data are: Job Type: CFO, CEO, Senior, vice president, manager

- 1. Degree: Doctoral, Bachelors, Masters, High School
- 2. Major: Math, Literature, Engineering, Business, Physics, Chemistry
- 3. YearsExperience:
- 4. Industry: Health, Service, Finance, Product, Web, Education
  - Miles from Metropolis:
- 5. Salary:

The calculations that will be performed for working of this proposed system to predict the salary with results:

Step 1: In step 1 we consider only Years' Experience vs Salary to create a base Model.

Here X is the independent variable which is the "years' Experience". And y is the dependent variable which is the "Salary".

Step 2:

- 1. Fit linear regression model to database
- 2. Firstly, building a simple Linear Regression model to see what prediction it makes.
- 3. We will be using the LinearRegression class from the library sklearn.linear\_model. We create an object of the LinearRegression class and call the fit method passing the X and y.



Figure -1: System architecture diagram





Chart -1: Base model result

- 2. So, we see here that the line created by our model is quite accurate.
- 3. The accuracy of our model is now around 96% to 98%.

To create a basic training model, two variables will be assigned for the model to use. Twenty percent of the training data will be split into testing data that we can use to test the model with data for which the salaries are already known.



# || Volume 6 || Issue 5 || May 2021 || ISSN (Online) 2456-0774 INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH AND ENGINEERING TRENDS

Step 4:

- 1. Now our main challenge is to add more parameters and maintain the accuracy
- 2. Next, we visualize each categorical (jobType, degree, major, industry, yearsExperience, milesFromMetroplis) feature to see which features could be good predictors of salary.
- 3. So, by visualizing each category we come to know that, yearsExperience yearsExperience has the highest correlation with salary. jobType also seems to be correlated with salary.

## Step 5:

- 1. Create baseline Model
- Baseline model is created on the dataset that contains all features using Linear Regression Where we will 80% data for training and 20% data to check our model.
- Mean squared error (MSE) will be evaluated now along with accuracy to evaluate the baseline model's performance.

In [59]: # get the mean square error

print ('Wean Squared Error of our prediction model is : ', mean squared error(sal train, sal predict))

Mean Squared Error of our prediction model 384.2694830873358

# Figure -2: Base model MSE result

After this stage the MSE is very high. It is 384. Now our aim is to reduce it.

So, to reduce MSE to less than 360. We can use:

- a. Apply Polynomial Transformation
- b. Use Ridge Regression
- c. Use Random Forest

From these 3 ways we are continuing with polynomial transformation which actually reduced the MSE. Step 6:

- 1. We will be using the Polynomial Features class from the sklearn.preprocessing library for this purpose. When we create an object of this class — we have to pass the degree parameter.
- 2. Let's begin by choosing degree as 2. Appy Polynomial Transformation.

If we have 2 labels named x1 and x2, then after applying polynomial transformation of degree 2 the new features will be:

- 1,  $x1^1$ ,  $x2^1$ ,  $x1^2$ ,  $x2^2$ , (x1 \* x2)
- 3. Fit and Transform the variables with 2<sup>nd</sup> order polynomial and then create a Linear Regression model on the new data.

- In [58]: # Fit and transform the variables with 2nd order polynomial
  pr = PolynomialFeatures(2)
  feat\_x\_train = pr.fit\_transform(x\_train)
  - feat\_x\_test\_pr = pr.fit\_transform(x\_test)
  - þr.
- In [71]: #Create a new model using Polynomial Transformation
  poly = LinearRegression()
  poly.fit(feat x train, sal train)

## Figure -3: Polynomial transformation result

Here model updated using polynomial transformation.

Step 7: Predict the MSE and accuracy of new model



MSE of testing data- 356.4817325350039

## Figure -4: MSE after applying Polynomial Transformation



Chart -2: Baseline model result after adding Polynomial Transformation



- 3. The new MSE is 357.
- 4. Linear regression with second order polynomial transformation gave best predictions with MSE of 357 and accuracy of 76%. This meets the goal of reducing MSE to below 360.

### IV CONCLUSIONS

In this paper we proposed a salary prediction system by using a linear regression algorithm with second order polynomial transformation. For the proper salary prediction, we found out most relevant 5 features. The result of the system is calculated by suitable algorithm by comparing it with another algorithms in terms of standard scores and curves like the classification accuracy, theF1score, the ROC curve, the Precision-Recallcurve etc. We compared algorithms only for the basic model which only two attributes. Moreover, we continued with basic model and found out the most appropriate method to add more attribute and with highest accuracy of 76%.

In future work, we would like add graphical user interface to system and try to save and reuse trained model.

### ACKNOWLEDGEMENT

We would like to express our sincere gratitude to several individuals and organizations for supporting me throughout my Project. First, I wish to express my sincere gratitude to my supervisor, Professor D. M. Lothe, for his enthusiasm, patience, insightful comments, helpful information, practical advice and unceasing ideas that have helped us tremendously at all times in our research. Without her support and guidance, this project would not have been possible. I could not have imagined having a better supervisor in our study.

### REFERENCES

- Susmita Ray," A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th -16th Feb 2019
- [2] Sananda Dutta, Airiddha Halder, Kousik Dasgupta," Design of a novel Prediction Engine for predicting suitable salary for a job" 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN).
- [3] 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)
- [4] 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)
- [5] Mangui Wu, Shunmin Shu," Top Management Salary, Stock Ratio and Firm Performance: A Comparative Study of State-owned and Private Listed Companies in China"