



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 7, Issue 3 - V7I3-2150)

Available online at: <https://www.ijariit.com>

Sophisticated way to predict incrementation of pay using Machine Learning

K. Rohith

rohithkosgi@gmail.com

The ICFAI Foundation for Higher Education,
Hyderabad, Telangana

M. Rishika Reddy

reddyrishika90@gmail.com

The ICFAI Foundation for Higher Education,
Hyderabad, Telangana

P. Harshavardhan

harshavardhanpoojarivari@gmail.com

The ICFAI Foundation for Higher Education,
Hyderabad, Telangana

V. Aashrith Surya

tinkustallion9@gmail.com

The ICFAI Foundation for Higher Education,
Hyderabad, Telangana

ABSTRACT

The goal of this paper is to predict the salary of an individual after a particular year. The graphical representation of predicting salary may be a process that aims to develop a computerized system to take care of all the daily work of salary growth graph in any field and may predict salary after a particular period of time. This application can take the database for the salary system from the firm and makes a graph through this information from the database. It'll check the salary fields then import a graph which helps to watch the graphical representation. then it can predict a particular period of time salary through the prediction algorithm. It also can be applied in another effective predictions also.

Keywords— Machine Learning, Linear Regression, Polynomial Regression

1. INTRODUCTION

A prediction is a belief about a future event. A prediction is sometimes, though not always, based upon knowledge or experience. Future events are not necessarily certain, thus confirming exact data about the future is in many cases impossible, a prediction may be useful to help in preparing plans about probable developments. In this paper the salary of an employee of an organization is to be predicted on the basis of previous salary growth rate. Here the history of salary has been observed and then on the basis of that salary of a person after a certain period of time it can be calculated automatically.

In this paper the main aim is predicting salary and making a suitable user-friendly graph. From this prediction the salary of an employee can be observed according to a particular field according to their qualifications. It helps to see the growth of any field. It can produce a person's salary by clustering and predict the salary through the graph. Applying linear regression

and polynomial regression it makes a graph. This graph helps to predict the salary for any position.

The application is aimed to develop to maintain a day-by-day monitoring to see the graphical medium of any field (salary or experiences as well as designation, etc.). A polynomial term is a quadratic (squared) or cubic (cubed) term turns a linear regression model into a curve. As the data X that is squared or cubed, not the Beta

coefficient, it still qualifies as a linear model. This makes it a simple and uncomplicated way to model curves without having to model complicated nonlinear models. One common pattern within machine learning is to use linear models trained on nonlinear functions of the data. This perspective maintains the generally fast performance of linear methods while allowing them to fit a much wider range of data. That helps for the curving design.

That will assist the employee as per following ways:

- Helping to see the growth in any field.
- With the assistance of machine learning it can effortlessly produce a graph.
- Marketing easy to estimate the salary between x-y axes.
- Users can give any point to get the salary through the program.
- Salary of the employees can be observed to give them a particular field according to their qualifications.
- The graphs through the Linear and polynomial graphs are displayed to detect the salary and position levels.

2. METHODOLOGY

Machine Learning (ML) [1] is fundamentally in the field of computer science with the assistance of which computer systems

can provide sense to data in much the same way as human beings do. In easy words, ML [2] [3] is a type of AI that withdraws patterns out of raw data by using an algorithm or method. The key focus of ML is to permit computer systems to learn from experience without being directly programmed or human interference.

Linear Regression [4] is an algorithm of machine learning based on a supervised learning scheme. Linear regression [6] carries out a task that may predict the value of a dependent variable (y) on the basis of an independent variable (x) that is given. Thus, this kind of regression approach looks for a linear type of connection between input x and output y.

Polynomial Regression [5] [7] is a form of linear regression in which the connection between the independent variable x and dependent variable y is modelled as an nth degree polynomial. Polynomial regression [8] fits a nonlinear connection between the value of x and the corresponding conditional mean of y, denoted $E(y|x)$.

This application provides a Salary graph representation that is mainly done by polynomial regression statistics, polynomial regression [8] is a form of regression analysis which represents the relationship between the independent variable x and the dependent variable y and that is modelled as the nth degree of polynomial in x. Polynomial regression is suitable for a nonlinear type of relationship between the value of x and the correlating conditional mean of y, represented as $E(y|x)$. Even though polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function $E(y|x)$ is linear in the unrevealed parameters that are estimated from the data. For this reason, polynomial regression is assumed to be a special case of multiple linear regressions.

Curve fitting is a method of building a curve, or to represent a mathematical function that is optimally suitable to a series of data points, and possibly it is subject to constraints. Curve fitting can require either interpolation, where an apt to the data is needed, or smoothing, in which a "smooth" function is created that is approximately suitable the data. A interconnected topic is regression analysis, which focuses more on query of statistical inference such as how much uncertainty is present in a curve that is fit to data observed with random errors. Fitted curves can be utilized as an aid for data visualization, to infer values of a function where no data are available, and to sum up the connections among two or more variables. Extrapolation refers to the apply of a fitted curve beyond the range of the noticed data, and is subject to a degree of unreliability as it may reflect the technique used to build the curve as much as it reflects the noticed data.

Following are the descriptions of the method that the work has been done. Proposed Method for Salary Prediction:

Step 1: Salary data have been taken from the dataset.

Step 2: Then the points corresponding to the salary data of an individual person have been plotted in the graph. The data are initialized in pandas (ascending, descending, mixed-up). Taking the dataset from each pandas field and from the pandas dataset we plotted the points on the graph as per number wise or input wise that came from the real dataset.

Step 3: After that we use linear regression to draw lines between the points.

Step 4: If the points are not in a linear way then we use polynomial regression for curving purposes. Through

the clustering points we can make a smooth and curved path.

Step 5: After then through the linear/polynomial graph through the x-y axis we can predict salary.

Step 6: Also, we predict a person's future salary position as per the graph. Only take a particular person's position, then the prediction answer will be executed through the help of the graph.

3. EXPERIMENTATION

This prediction has been implemented through the subsequent approaches:

- **numpy:-** NumPy may be a Python package which stands for 'Numerical Python'. It's the core library for scientific computing, which contains a strong n-dimensional array object, provide tools for integrating C, C++ etc. it's also useful in algebra, random number capability etc.
- **matplotlib.pyplot:-** matplotlib.pyplot is a collection of command style functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area during a figure, plots some lines during a plotting area, decorates the plot with labels, etc.
- **pandas:-** pandas is a kind of software library written for the Python programming language for data manipulation and analysis. It offers data structures and operations for manipulating numerical tables and statistics. Its free software released under the three-clause BSD license.
- **read_csv:-** Python may be a great language for doing data analysis, primarily due to the superb ecosystem of data-centric python packages. pandas is one among those packages and makes importing and analyzing data much easier.
- **Rectilinear regression:-** rectilinear regression is going to be wont to perform linear and polynomial regression and make predictions accordingly. Now, you've got two arrays: the input x and output y. you ought to call. reshape () on x because this array is required to be two-dimensional, or to be more precise, to possess one column and as many rows as necessary,
- **Polynomial features:-** Generate a replacement feature matrix consisting of all polynomial combinations of the features with degree but or adequate to the required degree. For instance, if an input sample is 2 dimensional and of the shape [a, b], the degree-2 polynomial features.
- **Predict:-** predict the values where found at the period of time that what proportion there's in that graph.

First, a dataset in excel file to be made then to be opened in Jupyter notebook from ANACONDA Navigator. From there we will read the dataset in ANACONDA Navigator. At Jupyter notebook first we take three variables for importing purposes like NumPy, matplotlib.pyplot and pandas. These functions are used are as follows:

- NumPy is employed for antialiasing a dynamic array or large set.
- matplotlib.pyplot is employed for creating graph
- pandas are mainly used like databases where we store the dataset from the excel file. Through the info set we are forming graph and predicting

Then importing the dataset on the pandas through the assistance of "read_csv" then showing that field that's imported from the excel file. The salary sets the plotting the points at the graph. After then it's being tried to input a linear regression function that creates only a line. If all points are linear then this function

is employed and directly goes making graphs and predicting. But if the points aren't linear, we choose the subsequent step. If the points aren't linear then we use a polynomial function for curving purposes that help to watch users how the expansion is moving on a path. After this the diagram on the graph is for demonstration. Then it's predicted to be a salary through the graph using both the x-y-axis. ".predict(poly_reg.fit_transform(v))".

From here it gets a prediction value through the polynomial/linear regression.

4. RESULTS AND DISCUSSION

This application is often observed because of the graphical representation and may also predict any point from position and automatically calculates salary. Also survey the salary venue.

Surveys of salary are therefore differentiated on basis of their data source into people who -

- * Get data from companies, or
- * Collect data from employees.

Survey operators assigned for salary strive to urge the foremost significant input file in every possible way. there's no thanks to decide which approach is correct. the primary possibility may assure large companies, whereas the second choice is especially for comparative smaller companies.

Output of accessing data set is as follows:

Out[2]:

	Position	Level	Salary
0	Business Analyst	1	45000
1	Junior Consultant	2	50000
2	Senior Consultant	3	60000
3	Manager	4	80000
4	Country Manager	5	110000
5	Region Manager	6	150000
6	Partner	7	200000
7	Senior Partner	8	300000
8	C-level	9	500000
9	CEO	10	1000000

Figure 1: Sample Dataset

First we linear function. "LinearRegression()" from this function makes a line through the points.

Output:

Linear Regression

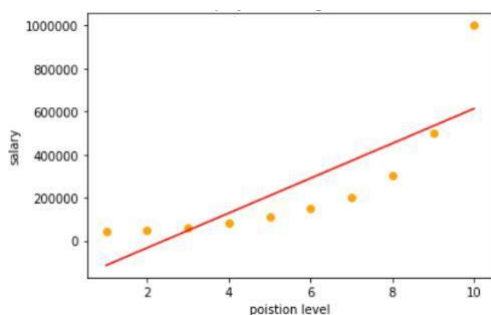


Figure 2: Results of rectilinear regression for the salary dataset

Now, plotting the purpose where x-axis represents the position and y-axis presents the salary from database Position_Salaries.

Output: Sampling

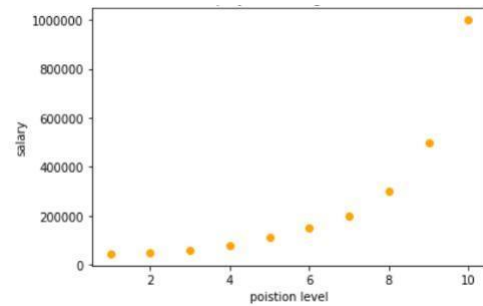


Figure 3: Sampling the info from the salary dataset

In Figure 3 we will see all the points aren't used. For that reason now we are implementing "PolynomialFeatures" for curveness. "degree = 6" here refers to the smoothness of the curve .

Same like above here, "lg2.predict(poly_reg.fit_transform(x))" function is employed for curving the line .

Output:

Polynomial Regression

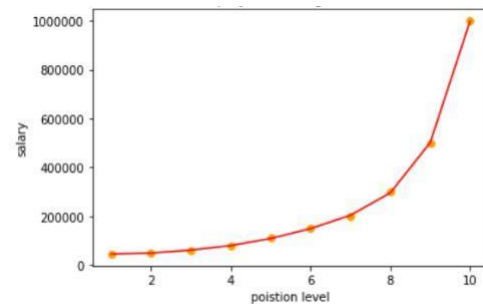


Figure 4: Results of polynomial regression for the salary dataset

5. CONCLUSIONS

Choosing the salary from x-y graph because it is used to represent a selected data set that takes some trial and error. Often there could also be quite one appropriate sort of graph to be used. It'll depend upon the way of selecting the thanks to present the info , also as its own preferences. Nowadays spreadsheet programs like Excel are very flexible to make graphs of various types; with a couple of clicks one can see the represented data as a bar chart also as by a line graph, or a circle graph. This prediction is correct upto a particular percentage. More accuracy is often obtained by implementing k-nearest regression. From there the simplest predictions are often chosen. It is often improved in following way:

- It can give more advanced software for tallying salary mediums.
- it'll host the platform on online servers.
- It may be as large as a database also and curve also bigger than the above example.

This predictor method is often used for predicting the population of a rustic also as forecasting a daily issue.

6. REFERENCES

[1] Andreas Muller, "Introduction to Machine Learning using Python: A guide for data Scientist," in O'Reilly Publisher, India.

- [2] S. Marsland, Machine learning: an algorithmic perspective. CRC press, 2015.
- [3] T.Zanis, George, et al “Modern Application of Machine Learning.”Proceedings of the 1st Annual SEERC Doctoral Student Conference-DSC.2006.
- [4] Mitchell, Tom Michael. The discipline of machine learning. Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2006.
- [5] Horvitz, Eric. “Machine learning, reasoning, and intelligence in daily life: Directions and challenges.” Proceedings of. Vol. 360.2006.