International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. **05** Issue **01** | **2020**

Employee Salary Prediction Using AWS Sage Maker

[1]Arpit Seth, [2] Gurubasava
[1]*MCA Scholar, School of CS & IT, Dept. of MCA, Jain (Deemed-to-be University) – 560069*
[2] *Assistant Professor, School of CS & IT, Dept. of MCA, Jain (Deemed-to-be University) - 560069*

ABSTRACT

The purpose of this paper is to estimate a person's salary after years of experience. The graphical representation of salary forecasting is a method aimed at creating a computerised system to retain all regular salary work based on the work experience in any sector and can estimate salary after a certain period of time. This prediction scheme will benefit the recruiting of businesses or H.R teams. The objective of this research is to predict the employee salary based on the number of years of experience.

1. INTRODUCTION

An inference regarding a potential event is a prediction. Often, but not always, a prediction is Based on knowledge or experience. Events in the future. They are not necessarily certain, thus accurately verified in several ways, information about the future is Impossible, and a forecast could be helpful to assist in the preparation of plans for likely developments. Advanced technology is the key to our lives. Almost every work uses the technology to satisfy the needs of society. Computer science is an important topic in the present age. It has a lot of real-life apps, such as cloud computing, artificial intelligence, remote surveillance, internet security, uncertainty, etc. Technology means that the use of information can be stored, retrieved, communicated and used [1]. So, the information can be conserved and shared with all organisations, industry and individuals using computer systems.

In all computer applications, internet security plays a major role. Suppose your company's HR department needs you to build a model for predicting the compensation of new workers on the basis of work experience. This research proposes to solve this problem with a simple linear regression. And it demonstrates that the model predicts well that the HR department will make future decisions to forecast salaries based upon the applicants' experience. The final outcome of this research is to come up with a linear regression model that could accurately predict employee salary based on number of years of experience.

In this paper the dataset is divided into years of experience as an input set and salary is predicted as output set. In this dataset years of experience is considered as independent variable and salary is dependent variable. These factors are directly proportion to each other. In simple linear regression, we predict the value of one variable Y based on another variable X. X is called the independent variable and Y is called the dependant variable.

The hybrid machine learning and cloud computing architecture offers us the ability to keep the data as well as a model that can predict employee salaries. The definition here is clear and incredibly simplistic and this is not the definition that would be addressed in other people. Amazon Sage Maker is a fully-managed service that enables every developer and data scientist to easily design, train and deploy ML models. In order to promote the creation of high-quality models, Sage Maker eliminates the heightened lifting from every stage of the learning process. Amazon Sage Maker helps in training models with different endpoints. After model testing, the model with highest accuracy cam be used.

1.1. Motivation

This research is helpful to both the employee and an organization. There was no transparency on the criteria that how packages are offered to the employees. And this research will also be helpful for organisations so that companies can offer efficient and appropriate packages according to business point of view.

2. LITERATURE REVIEW

This article focuses on how machine learning and cloud computing contribute to successful management of workers and why machine training in organisations is needed to classify the wages on the basis of experience. [2] Ignacio and Mariello methodology is based on multi-characteristic generalisation of the MI. The MI is

evaluated in particular between the selected characteristics and the group, to add only those characteristics which are relevant when taken together.

Sriramakrishnan Chandrasekara [3] proposed that the CSV data can save the data in the same file on an ongoing basis. This needs some cloud backup. The files stored in AWS will be used to collect and update data continuously in the model. Pornthep Khongchai in this [4] paper offers a wage prediction method using a student profile as a model. A technique for data mining is used to build a model for predicting salaries for students with similar qualities to the data. In this study, we have also experimented with the comparison of five techniques for data mining, including Decision trees, Naive Bayes, neighbour K-Nearest, Vector support machines, and neural networks, to find effective methods for wage prediction.

3. DATASET DESCRIPTION

Dataset used in this project is classified in two column that is year of experience and salary. For the formula generation we suppose salary as 'X' and year of experience as 'Y'. We predict the value of one variable 'Y' based on another variable 'X'. X is called the independent variable and Y is called the dependant variable. When the variable separately increases (or decreases), the relying variable linearly increases (or decreases). The data is collected from different sources. All the data will be stored in AWS S3 storage. And all the testing and training is performed on the dataset in Sage Maker instances.

Table: Dataset

INDEX	YEAR OF EXPERENCE	SALARY
1	1.1	39343
2	1.3	46205
3	1.5	37731
4	2	43525
5	2.2	39891
6	2.9	56642
7	3	60150

4. PROPOSED METHODOLOGY

We need computer education in the corporate sector to develop their company and to recognise the employee's package to obtain their expertise. This segment focuses on the benefits of machine learning and cloud computer in the corporate sector and on how we can effectively access these things in order to design and apply our prediction models.

Usage of machine training in the business sector, we carry on machine learning initiatives for our business growth in an organization and recruit people to do a good job. But here, we need to tackle our own business with machine learning so that we can recognize real things from our departments. The key explanation is for enhancing hiring, management of wages and management of workers. Here we have to define what the variables for the prediction model are as follows Number of years of experience and Package

The number of years of experience indicates the need for the employee's package in the business. How many years he served in the related area of recruiting is the basis of this criterion. Package is the dependent variable and on the basis of our independent years of experience, we will forecast this component. In the next model design, we shall define additional elements for variables that could be defined as independent variables that could serve as the most significant independent variables for the design and implementation of better prediction models.

Data are gathered in two areas: one is an independent variable such as years of experience and a dependent variable such as pay. Machine learning algorithms such as simple linear regression are applied here and we discuss the same in the current system section. Another issue is why this proposed architecture includes cloud computing and which specifications are defined here in lateral sections in the architecture for implementing cloud computing. In the next section we will be discussing the machine-learning algorithms we have used with simple algorithms, the parquets which we are using and the artefacts we implement for the p. The main purpose of this article is to define the technologies we use in current methods and to demonstrate the difference between existing and proposed arches.

"Numpy" is the library for mathematical operations that we use here for the simple, slight regression, when we recognize the pitch of the variables that the decision boundary is drawn between the points of the results. We are performing a simple linear regression method and the issue here is the recognition of the variable slope. We need to train the model with training data and test information to assess the employee's expected salary on the basis of our experience, which is anticipated. X defines the variable independent and Y defines the wage dependent variable. In the image below we will describe the points where the prediction will be carried out as the data points.

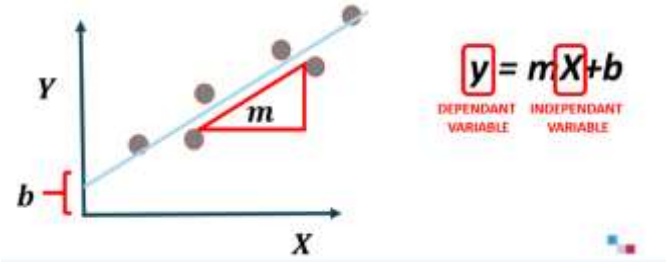

Fig – formula

Once the coefficients m and b are obtained, you have obtained a simple linear regression model! This "trained" model can be later used to predict any salary based on the number of years of experience. Using scientists, we will use our prediction model with regard to the implementation principle. In this respect, the slope of the variables will be determined based on a simple linear regression formula to define the decision boundary. The limit of the simple linear regression will define the new node group, which we shall see as the new entry. We have a common colour of nodes since only one independent variable is implemented by this basic linear regression. But we need to find out the group and the node field to be applied to the cluster as new data.

Fig – Relationship between parameters

We divide the data into a format of 20-80, which takes into account the training data and trains the model with that data. If any test data is in the package, it enters the new role based on the experience.

4.1 Machine Learning and Cloud Computing Implementation in Corporate

In any area, machine learning and cloud computing are needed for data storage and data processing and for data management. We offer AWS services, which we offer for data storage and administration, as part of cloud computing. The CSV data would save the data in the same file continuously. This needs some cloud backup. The files stored in AWS are used to constantly collect and update data for the model we use. The following is the AWS EC2 and S3 architecture, which helps us to store data and map data to the IAM service model.
AWS Sage maker helps in training a prediction model with the help of build in algorithms with can be used directly in AWS sage maker.

First, we need to import sage maker and boto 3 in notebook. Boto3 is the Amazon Web Services (AWS) Software Development Kit (SDK) for Python. Boto3 allows Python developer to write software that makes use of services like Amazon S3 and Amazon EC2. Now, we need to create a Sage maker session. Sage maker helps in storing the data or training dataset in s3 and after the model is trained it is also stored in S3 so that the model can be used later. The main advantage of sage maker is that it trains N number of models out of which the best or most accurate model can be used. Linear Learner is a supervised learning algorithm that is used to fit a line to the training data.

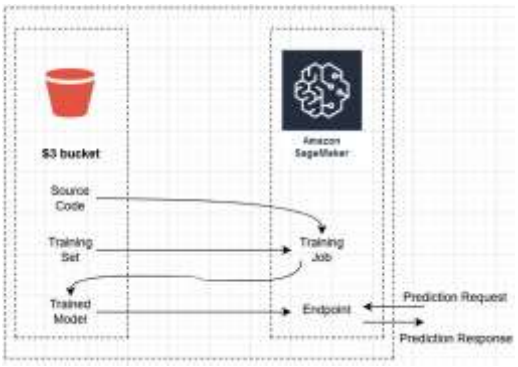

Fig - Flow chart

5. RESULT

This is the expected outcome. The findings and the sample codes that we use in this approach will be demonstrated in this section. In this architecture we used linear regression model that could accurately predict employee salary based on number of years of experience. Least squares fitting is a way to find the best fit curve or line for a set of points. The sum of the squares of the offsets (residuals) are used to estimate the best fit curve or line. Least squares method is used to obtain the coefficients m and b.

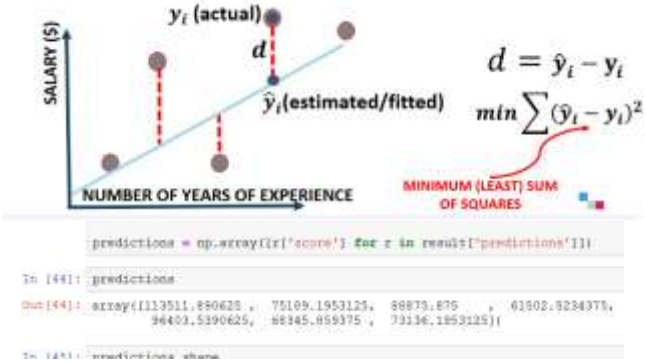

Fig- Prediction from the trained model

5.1 Visualize Test Set Results

This creates a linear straight-line relationship between salary and work experience.

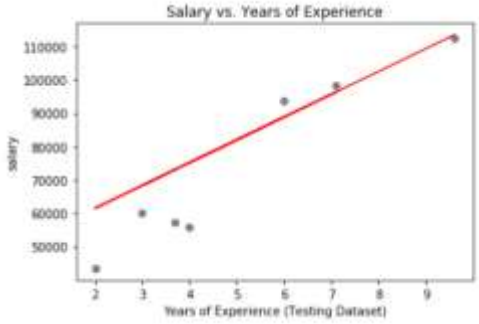

Fig-Test Set Results

6. CONCLUTION

In this architectural approach, we concentrate on the idea of defining models to enhance the execution of pay models using the real-time scenario and to efficiently improve the financial affairs of the company. We have one dependent and one independent variable in the basic linear regression and that will establish a simple relationship that supports the estimation of wages on the basis of our work experience. According to the results the proposed model overcome some drawbacks. The future scope of improvement includes: There is no user interface provided in this project. Other factors can be used for more effective prediction.

7. REFERENCES

[1] Seth, "GENRE PREDICTION FOR MUSIC RECOMMENDATION USING MACHINE LEARNING," EPRA International Journal of Research and Development (IJRD), vol. 5, no. 4, pp. 206-210, 2020.

[2] M. A. a. B. R. a. H. Ignacio, "Salary Prediction in the IT Job Market with Few High-Dimensional Samples: A Spanish Case Study," International Journal of Computational Intelligence Systems, vol. 11, pp. 1192-1209, 2018.

[3] S. Chandrasekaran, "A Machine Learning Implementation of Predicting the Real Time Scenarios in a better way," International Journal of Pure and Applied Mathemati, vol. 119, pp. 1301-13011, 2018.

[4] P. K. a. P. Songmuang, "2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)," in Implement of salary prediction system to improve student motivation using data mining technique, 2016, pp. 1-6.