**Machine Learning**

**Project Name:** Salary prediction using linear regression analysis

**Group Number:** 01

**Group Members:**

* Afrida Rahman ( 19-39557-1 )
* Samir Faisal ( 19-41037-2 )
* Hasan Kazi Mahmud ( 19-41089-2 )
* Shazzad Hossain Shovon ( 17-34884-2 )

**Introduction**

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable. A regression problem is when the output variable is a real or continuous value, such as “salary” or “weight”. Many different models can be used, the simplest is the linear regression. It tries to fit data with the best hyper-plane which goes through the points. It is also one of the limitations of linear regression. Outlier: An outlier is an unusual observation of response y, for some given predictor x. High Leverage Points: Contrast to an outlier, a high leverage point is defined as an unusual observation of predictor x. A linear regression line has an equation of the form Y = a + bX, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when x = 0).

Problem:Finding the closest salary value from a large dataset is quite impossible for human being.

Solution: The linear regression is one of the simplest algorithm which can compete with the most accurate models because it makes highly accurate predictions.

**State-of-the-arts**

|  |  |  |  |
| --- | --- | --- | --- |
| Research | Year | Techniques | No of Datasets |
| Mahdi Pakdaman Naeini  Hamidreza Taremian | 2010 | neural networks | 2000 |
| D C K I WILLIAMS | 1997 | linear regression, using Gaussian process | 3000 |

**Proposed Method**

After loading 100 Dataset of year and salary, we will be predicting the dataset model. Then load the test value and predict. After that we will predict salary with the dataset and find the efficient solution.

Dataset select

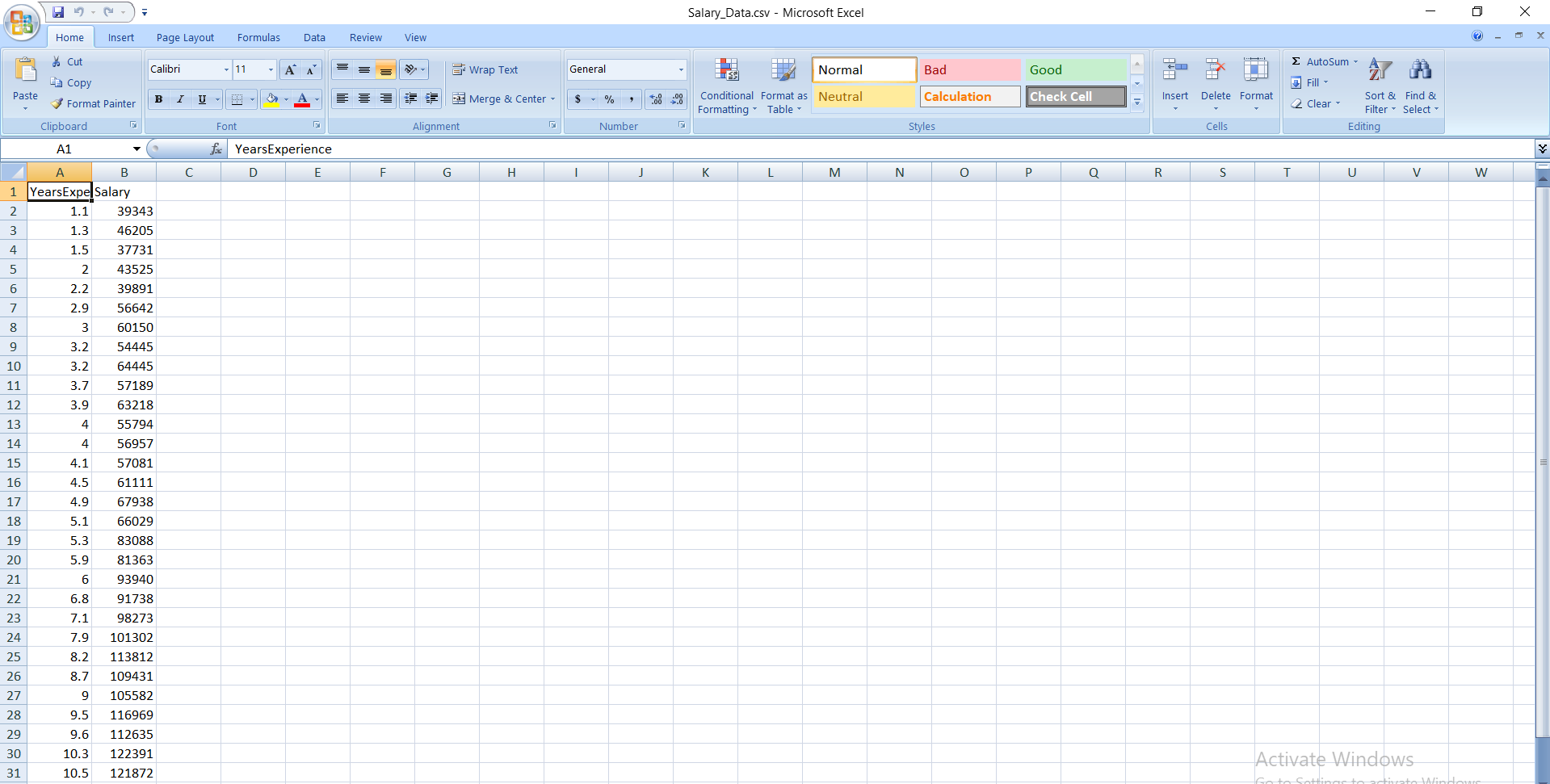
Predict salary value

Dataset testing

Training half of dataset

**DATASETS**

Dataset: we will download the dataset from https://www.kaggle.com/datasets/shubham47/salary-data-dataset-for-linear-regression?resource=download website where they stored the dataset in different versions, python, Matlab and binary. There are 1 files contains 31 data from salary and year.



**Analysis**

Predict the salary dataset helps to determine close value for dataset using Linear Algebra setting gradient equal to zero formula(w = (xTx)-1 xTy. For that both salary and year we will use different dimension(x,y).  Training our half of the dataset using linear algebra and then we get a model from the algorithm, then we will test the dataset using this model and we will predict efficient salary value.

**Discussion**

We will predict salary from dataset salary and year using linear regression algorithm. For nearest dataset value of slope (y=mx+c) to find using linear algebra.

To determine the salary prediction depends on job experience from the model which is based on linear regression model.

**Reference**

1. <https://study.com/academy/lesson/problem-solving-using-linear-regression-steps-examples-quiz.html>
2. <https://www.kaggle.com/datasets/shubham47/salary-data-dataset-for-linear-regression?resource=download>
3. <https://www.google.com/search?q=linear+regression&source=lmns&hl=en&sa=X&ved=2ahUKEwjUxJOpp6z4AhUoyKACHf3VDz8Q_AUoAHoECAEQAA>
4. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.1226&rep=rep1&type=pdf>