

# Predicting Company Profit

## Data Science

---

### Introduction

A dataset with R&D Spend, Administration Costs and Marketing Spend of 50 Companies are given along with the profit earned. The project aims to explore and compare various machine learning regression algorithms to identify the best model for accurate profit prediction.

### Methodology

- **Exploratory Data Analysis and Preprocessing**

The dataset was checked for any missing values and outliers. There aren't any. All the variables are continuous. Histograms and a heat map are plotted to observe the data. We can observe that the R&D Spend and the company's profit are highly correlated.

- **Data Splitting**

The dataset was divided into training and testing sets to evaluate the performance of the regression models. An 80-20 split was used, with the larger portion dedicated to training data.

---

---

- **Regression Algorithms**

Two regression algorithms were used. They are linear regression and ridge regression with alpha value 1.

- **Model Evaluation**

The following regression metrics were used to evaluate the performance of each model:

- R-squared(R<sup>2</sup>) score
- Mean absolute error
- Mean squared error
- Mean squared log error

## **Conclusion**

Based on the regression metrics and overall performance, linear regression was the best algorithm for this specific profit prediction task.

## **Libraries and their versions**

- Numpy: 1.24.3
- Pandas: 2.0.1
- Seaborn: 0.12
- Matplotlib: 3.5.1
- Scikit-learn: 1.2.2