

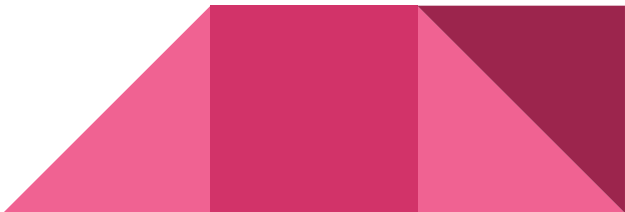


T.Raghavendra Prasad, Technical Officer 'B'

Combat Aircraft Systems Development and Integration Centre

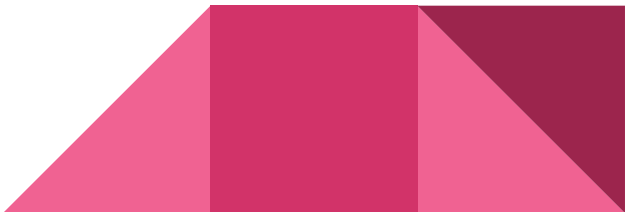
Defence Research and Development Organisation

Using Multiple Linear Regression to Predict Profits for Startups

- Machine Learning Model is used for predicting the profit from the startup's dataset with the given features.
 - We will use the 50-startups dataset for this problem statement and we will be using the concept of Multiple linear regression to predict the profit of startups companies.
- 

How this project helps the startup?

Startups are typically funded through various sources and methods, such as community-based funding, crowdfunding, angel investors, venture capitalists, and government grants/programs, among others. The funding sources and methods used by startups are similar to those used by other multinational corporations (MNCs), but the main difference is that startups work on a small scale to create innovative products that benefit customers, while established companies work on a larger scale by improving existing products.



How Startups are funded?

Startups are often funded through seed funding, which involves early-stage investors providing capital to help turn an idea into a product. This type of funding is typically provided by angel investors or venture capitalists who are willing to take on the risk of investing in a company that does not yet have a proven track record or market value.




How the Linear Regression Model helps?

This particular machine learning model can be beneficial in situations where there is a need to determine profitability based on market expenditures. Essentially, the model can assist in identifying profits based on the amount of money spent from the 50 startup dataset.



About the Startup Dataset:

This dataset contains information on 50 startups located in New York, California, and Florida. The dataset includes data on R&D spending, administration spending, marketing spending, and location features, with the target variable being profit. The dataset provides information on how much each startup spends on research and development, administration, and marketing, as well as the state in which each startup is located, and the amount of profit each startup generates.




Difference between Linear regression and Multiple linear regression

Primary distinction between linear regression and multiple linear regression is that linear regression involves a single independent variable, whereas multiple linear regression involves more than one independent variable.

To illustrate, consider the following examples:

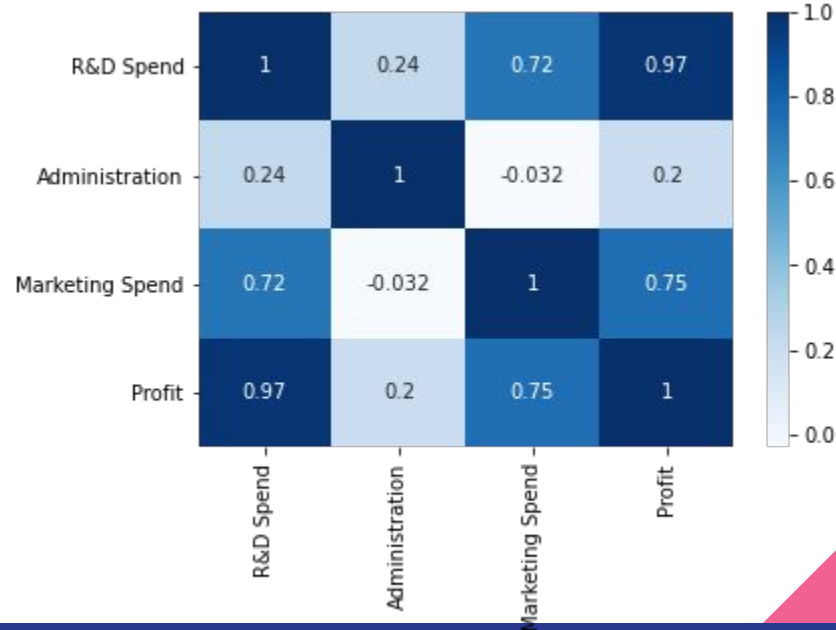
Linear regression: Predicting an individual's height based on their weight.

Multiple linear regression: Modifying the above scenario slightly by including additional features such as age, gender, and height to predict an individual's weight, which would require the use of multiple linear regression.



Correlation between the columns


To find out the correlation between the independent variables R&D Spend, Administration and Marketing Spend against the dependent variable profit we use the `corr()` function in pandas.



Conclusion

| Model Evaluation Parameters | Value |
|-----------------------------|-------------------|
| R Squared Score | 93.39448007716636 |
| Mean Squared Error | 6224496238.94644 |
| Root Mean Squared Error: | 788954.7666974603 |
| Mean Absolute Error: | 6503.577323580025 |

For the above Model Evaluation parameters, we can conclude that since R Squared score is 93.94%, the Linear Regression Line is best fit for the given data set and Mean Absolute Error of 6503.577. Therefore our predicted value can be 6503.577323580025 units greater than or less than the actual value.



References

- <http://www.ltcconline.net/greenl/courses/201/regression/scatter.htm>
- <https://www.statisticshowto.com/probability-and-statistics/regression-analysis/find-a-linear-regression-equation/>
- <https://www.statology.org/scatterplot-with-regression-line-python/>
- <https://www.kaggle.com/datasets/farhanmd29/50-startups>

