DSP PROJECT

NAME : KALAGATHA KEERTHI

ID NO : S180375

Project name : PROFIT PREDICTION

**DESCRIPTION OF THE PROJECT :**

*The dataset that I am using for the task of profit prediction includes data about the R&D spend, Administration cost, Marketing Spend, State of operation, and the historical profit generated by 50 startups.*

*The profit earned by a company for a particular period depends on several factors like how much time and money a company spends on R&D, marketing and many more. So for predicting the profit of a company for a particular period we need to train a machine learning model with a dataset that contains historical data about the profit generated by the company*.

*The task of predicting profit is an important task for every business to set an achievable goal. For example, if the business spends $500 on marketing, it can’t expect a profit of $20,000. Likewise, there are many other factors on which the profit of a business depends. A company must therefore set a goal that can be achieved.*

TOOLS USED FOR PROJECT:

**1.PANDAS** : **pandas** is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language

**2.NUMPY**: *Numpy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.*

3**.MATPLOTLIB**: *Matplotlib is a python library* ***used to create 2D graphs and plots by using python scripts****. It has a module named pyplot which makes things easy for plotting by providing feature to control line styles, font properties, formatting axes etc*.

**4.SEABORN**: *Seaborn is* ***an open-source Python library built on top of matplotlib****. It is used for data visualization and exploratory data analysis. Seaborn works easily with dataframes and the Pandas library. The graphs created can also be customized easily.*

5.SKLEARN : *Scikit-learn (Sklearn) is the most useful and robust library for* ***machine learning*** *in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python*

**6.LINEAR REGRESSION:*Linear Regression*** *is a machine learning algorithm based on* ***supervised learning****. It performs a* ***regression task****. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables getting used.*

***INPUT*** *: R&D spend, Administration cost, Marketing Spend, State of operation,profit*

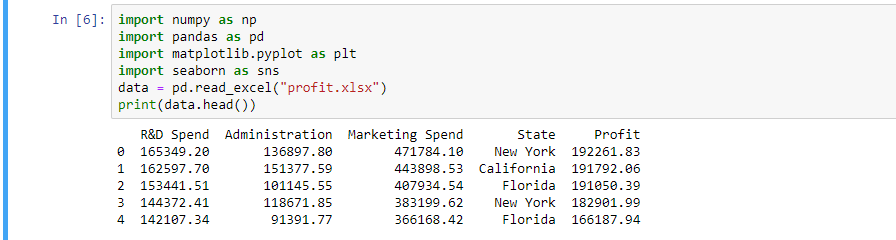
**OUTPUT** : *predicted*

*Git hub link:* [*https://github.com/keerthik1911/dsp-project.git*](https://github.com/keerthik1911/dsp-project.git)

*.*

PROFIT PREDICTION USING PYTHON:

#importing the necessary Python libraries and the dataset:



**DESCRIPTION OF THE DATASET**:

