Linear regression model:

The linear regression model uses one independent variable to predict a dependent variable (target)

Here the independent variable is R&D spend

Dependent variable is profit

Now we are going to use this IDV to predict the DV

First, we split the data for testing and training

We give test size = 0.2, that means that 20% of the data is for testing and rest 80% is given for training the model

Then after running the testing and training stages of model , we calculate the r2 score of model , it gives the explanatory power of the model.

Here r2 score = 0.9265, that means that 92.65% of the changes in the target variable profit can be explained by the independent variable R&D spend, the rest of the % of changes that can’t be explained is intercept i.e unexplainable part. Here the intercept is less and r2 score is high, so it’s a good prediction model.

Here intercept = 49336 (unexplainable part – the lesser the better model is)

Therefore, based on the details we can formulate the linear regression model for this,

Predicted variable – profit = intercept + (coefficient of R&D parameter) \* (R&D value)

Multiple regression model 1

Here we take all the 4 given attributes in the dataset as independent variables to predict the dependent variable profit

We get R² Score: 0.9001, i.e about 90% of changes in the profit can be explained by the 4 IDVs - R&D Spend, Administration, Marketing Spend, State.

Here intercept (unexplainable changes) = 54080 (this is higher then in previous model due to lesser explanatory power of model evident by r2 score)

We try to improve this regression model by dropping some IDVs

Multiple regression model 2

Here we take 2 IDVs - R&D Spend and Marketing Spend

DV – profit

R2 score = 0.9168, i.e 91.68% of changes in profit can be explained by the 2 IDVs, thus this model has higher explanatory power than the previous regression model but still lesser than the linear regression model

Hence, here the best model for accurate prediction of profit is the linear regression model with IDV as R&D spent.