**To find the best suitable model for the given Problem Statement and to predict profitability**

**Dataset of the statement**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **R&D Spend** | **Administration** | **Marketing Spend** | **State** | **Profit** |
| 165349.2 | 136897.8 | 471784.1 | New York | 192261.8 |
| 162597.7 | 151377.59 | 443898.53 | California | 191792.1 |
| 153441.51 | 101145.55 | 407934.54 | Florida | 191050.4 |
| 144372.41 | 118671.85 | 383199.62 | New York | 182902 |
| 142107.34 | 91391.77 | 366168.42 | Florida | 166187.9 |
| 131876.9 | 99814.71 | 362861.36 | New York | 156991.1 |
| 134615.46 | 147198.87 | 127716.82 | California | 156122.5 |
| 130298.13 | 145530.06 | 323876.68 | Florida | 155752.6 |
| 120542.52 | 148718.95 | 311613.29 | New York | 152211.8 |
| 123334.88 | 108679.17 | 304981.62 | California | 149760 |
| 101913.08 | 110594.11 | 229160.95 | Florida | 146122 |
| 100671.96 | 91790.61 | 249744.55 | California | 144259.4 |
| 93863.75 | 127320.38 | 249839.44 | Florida | 141585.5 |
| 91992.39 | 135495.07 | 252664.93 | California | 134307.4 |
| 119943.24 | 156547.42 | 256512.92 | Florida | 132602.7 |
| 114523.61 | 122616.84 | 261776.23 | New York | 129917 |
| 78013.11 | 121597.55 | 264346.06 | California | 126992.9 |
| 94657.16 | 145077.58 | 282574.31 | New York | 125370.4 |
| 91749.16 | 114175.79 | 294919.57 | Florida | 124266.9 |
| 86419.7 | 153514.11 | 0 | New York | 122776.9 |
| 76253.86 | 113867.3 | 298664.47 | California | 118474 |
| 78389.47 | 153773.43 | 299737.29 | New York | 111313 |
| 73994.56 | 122782.75 | 303319.26 | Florida | 110352.3 |
| 67532.53 | 105751.03 | 304768.73 | Florida | 108734 |
| 77044.01 | 99281.34 | 140574.81 | New York | 108552 |
| 64664.71 | 139553.16 | 137962.62 | California | 107404.3 |
| 75328.87 | 144135.98 | 134050.07 | Florida | 105733.5 |
| 72107.6 | 127864.55 | 353183.81 | New York | 105008.3 |
| 66051.52 | 182645.56 | 118148.2 | Florida | 103282.4 |
| 65605.48 | 153032.06 | 107138.38 | New York | 101004.6 |
| 61994.48 | 115641.28 | 91131.24 | Florida | 99937.59 |
| 61136.38 | 152701.92 | 88218.23 | New York | 97483.56 |
| 63408.86 | 129219.61 | 46085.25 | California | 97427.84 |
| 55493.95 | 103057.49 | 214634.81 | Florida | 96778.92 |
| 46426.07 | 157693.92 | 210797.67 | California | 96712.8 |
| 46014.02 | 85047.44 | 205517.64 | New York | 96479.51 |
| 28663.76 | 127056.21 | 201126.82 | Florida | 90708.19 |
| 44069.95 | 51283.14 | 197029.42 | California | 89949.14 |
| 20229.59 | 65947.93 | 185265.1 | New York | 81229.06 |
| 38558.51 | 82982.09 | 174999.3 | California | 81005.76 |
| 28754.33 | 118546.05 | 172795.67 | California | 78239.91 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 27892.92 | 84710.77 | 164470.71 | Florida | 77798.83 |
| 23640.93 | 96189.63 | 148001.11 | California | 71498.49 |
| 15505.73 | 127382.3 | 35534.17 | New York | 69758.98 |
| 22177.74 | 154806.14 | 28334.72 | California | 65200.33 |
| 1000.23 | 124153.04 | 1903.93 | New York | 64926.08 |
| 1315.46 | 115816.21 | 297114.46 | Florida | 49490.75 |
| 0 | 135426.92 | 0 | California | 42559.73 |
| 542.05 | 51743.15 | 0 | New York | 35673.41 |
| 0 | 116983.8 | 45173.06 | California | 14681.4 |

|  |  |
| --- | --- |
| **Model** | **r\_score** |
| **Multiple linear regression** | 0.9358680970046243 |

**Support vector machine**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **s.no** | **Hyper parameter** | **Linear** | **Rbf** | **Poly** | **Sigmoid** |
| 1 | **C=0.1** | **0.9375216516281204** | -0.057469387821565965 | -0.056824517369874705 | -0.05748758102694351 |
| 2 | **C = 1 (default value)** | 0.8950779235664468 | -0.05731730927224388 | -0.050890117824376135 | -0.0574991971677592 |
| 3 | **C = 10** | -2.4372150243234123 | -0.055800922934202024 | 0.02531238887543097 | -0.05761538606317651 |
| 4 | **C = 100** | -357.07951114177723 | -0.03023555979437731 | 0.46566263381175776 | -0.05878002374292657 |
| 5 | **C= 500** | **Increasing in negative value** | 0.050018181433489683 | 0.620773805058096 | -0.06401665700120085 |
| 6 | **C = 1000** | 0.16060029222433436 | 0.6403239377679872 | -0.070701273098142 |
| 7 | **C = 2000** | 0.2883954414009646 | 0.6717477146409396 | -0.08453325370568487 |

**The support vector machine's highest r\_score value is 0.9375216516281204 using a linear hyperparameter C=0.1.**

**Decision Tree**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **s.no** | **Criterion** | **Splitter** | **max\_features** | **r\_score** |
| 1 | **Squared\_Error** | **best** | **auto** | 0.611573845 |
| 2 | **sqrt** | 0.538276136 |
| 3 | **log2** | 0.436129483 |
| 4 | **random** | **auto** | 0.683968153 |
| 5 | **sqrt** | 0.508036926 |
| 6 | **log2** | 0.378445582 |
| 7 | **Friedman\_Mse** | **best** | **auto** | 0.521732947 |
| 8 | **sqrt** | -0.16855781 |
| 9 | **log2** | 0.7119046 |
| 10 | **random** | **auto** | 0.520327763 |
| 11 | **sqrt** | 0.051820863 |
| 12 | **log2** | 0.257659282 |
| 13 | **Absolute\_Error** | **best** | **auto** | 0.606406558 |
| 14 | **sqrt** | 0.157994284 |
| 15 | **log2** | 0.529032249 |
| 16 | **random** | **auto** | 0.481981111 |
| 17 | **sqrt** | 0.173347969 |
| 18 | **log2** | 0.103136995 |
| 19 | **Poisson** | **best** | **auto** | 0.562251011 |
| 20 | **sqrt** | 0.538079847 |
| 21 | **log2** | 0.232578341 |
| 22 | **random** | **auto** | 0.546943685 |
| 23 | **sqrt** | 0.521352737 |
| 24 | **log2** | 0.60677934 |

* **The Decision Tree's highest r\_score value is 0.683968153** **using hyperparameter Criterion = Squared\_Error , Splitter= random, max\_features=auto**
* **• The r\_score value for the same hyperparameter was fluctuating constantly. Once more running the programme with the same hyperparameter**

**Random Forest**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **s.no** | **Criterion** | **N\_Estimators** | **Max\_Features** | **r\_score** |
| 1 | **Squared\_Error** | **10** | **sqrt** | 0.519141672 |
| 2 | **log2** | 0.519141672 |
| 3 | **auto** | 0.925277279 |
| 4 | **50** | **sqrt** | 0.683002237 |
| 5 | **log2** | 0.683002237 |
| 6 | **auto** | 0.944633639 |
| 7 | **100** | **sqrt** | 0.75915045 |
| 8 | **log2** | 0.75915045 |
| 9 | **auto** | 0.946004355 |
| 10 | **Absolute\_Error** | **10** | **sqrt** | 0.721083996 |
| 11 | **log2** | 0.721083996 |
| 12 | **auto** | 0.928182284 |
| 13 | **50** | **sqrt** | 0.722235187 |
| 14 | **log2** | 0.722235187 |
| 15 | **auto** | 0.940193525 |
| 16 | **100** | **sqrt** | 0.785748335 |
| 17 | **log2** | 0.785748335 |
| 18 | **auto** | 0.945909746 |
| 19 | **Friedman\_Mse** | **10** | **sqrt** | 0.527283 |
| 20 | **log2** | 0.527283 |
| 21 | **auto** | 0.920668118 |
| 22 | **50** | **sqrt** | 0.688918213 |
| 23 | **log2** | 0.688918213 |
| 24 | **auto** | 0.938895763 |
| 25 | **100** | **sqrt** | 0.760859221 |
| 26 | **log2** | 0.760859221 |
| 27 | **auto** | 0.941270197 |
| 28 | **Poisson** | **10** | **sqrt** | 0.752059569 |
| 29 | **log2** | 0.752059569 |
| 30 | **auto** | 0.930486613 |
| 31 | **50** | **sqrt** | 0.720862467 |
| 32 | **log2** | 0.720862467 |
| 33 | **auto** | 0.946354971 |
| 34 | **100** | **sqrt** | 0.771764207 |
| 35 | **log2** | 0.771764207 |
| 36 | **auto** | 0.941388942 |

* **The Random forest's highest r\_score value is 0.946354971 using hyperparameter Criterion = Poisson, n\_estimators=50, max\_features=auto**

**Conclusion**

|  |  |  |
| --- | --- | --- |
| **S.No** | **Model** | **r\_score** |
| 1 | multiple linear regression | 0.9358680970046243 |
| 2 | support vector machine | 0.9375216516281204 |
| 3 | decision tree | 0.683968153 |
| **4** | **random forest** | **0.946354971** |

As a result, **random fores**t is the finalised and has the **greatest r\_score value**.