**Insurance Charges Prediction**

1. End goal is Insurance Charge prediction.
2. 1338 Rows and 6 Columns.
3. Nominal Data- String convert to Numbers.
4. R2 researched value given below,

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **R2 values:** | | | | | | | | | | | |
|  |  | | |  | |  |  | |  |  | |
|  |  | | |  | |  |  | |  |  | |
| **1.AdaBoostRegression:** | | | | | | | | | | | |
|  |  | | |  | |  |  | |  |  | |
| **SL.NO.** | n-estimators | | | **linear** | | **square** | **exponential** | |
| 1 | **1** | | | **0.880957** | | **0.880957** | **0.880957** | |
| 2 | 50 | | | 0.844747 | | 0.518532 | 0.629285 | |
| 3 | 100 | | | 0.844747 | | 0.518532 | 0.538576 | |
|  |  | | |  | |  |  | |  |  | |
| **AdaBoostRegression:R2 value=0.880957 (Linear, n\_estimators=1)** | | | | | | | | |  |  | |
|  |  | | |  | |  |  | |  |  | |
| **2.xgboostRegression:** | | | | | | |  | |  |  | |
|  |  | | |  | |  |  | |  |  | |
| **SL.NO** | **n-estimators** | | | **Max\_depth** | | **R2 VALUES** |  | |  |  | |
| 1 | **100** | | | ***2*** | | **0.892045** |  |  | | |
| 2 | *100* | | | *7* | | 0.848112 |  | |  |  | |
| 3 | *1000* | | | *2* | | 0.866034 |  | |  |  | |
| 4 | *1000* | | | *7* | | 0.816192 |  | |  |  | |
|  | |  |  | |
|  | |  |  | |
|  | |  |  | |
|  | |  |  | |
|  |  | | |  | |  |  | |  |  | |
| **xgboostRegression - R2 value=0.892045 (n\_etimators-100,Max\_depth-2)** | | | | | | | | |  |  | |
|  |  | | |  | |  |  | |  |  | |
|  |  | | |  | |  |  | |  |  | |
| **3.LGBoostingRegression:** | | | | | | |  | |  |  | |
|  |  | | |  | |  |  | |  |  | |
| **SL.NO** | **n\_samples** | | | **n\_features** | | **R2 VALUES** |  | |  |  | |
| 1 | ***1000*** | | | ***20*** | | 0.866031 |  |  | | |
| 2 | ***1000*** | | | ***10*** | | 0.866031 |  | |  |  | |
| 3 | ***100*** | | | ***20*** | | 0.866031 |  | |  |  | |
| 4 | ***100*** | | | ***10*** | | 0.866031 |  | |  |  | |
|  | |  |  | |
|  | |  |  | |
|  | |  |  | |
|  | |  |  | |
|  |  | | |  | |  |  | |  |  | |
| **LGBoostingRegression - R2 value=0.866031(n\_samples=1000, n\_features-20)** | | | | | | | | | |  | |

Final model is **xgboostRegression - R2 value=0.892045 (n\_etimators-100,Max\_depth-2)**

**Final Output:**

**independent=dataset[['age', 'bmi', 'children','sex\_male', 'smoker\_yes']]**

**I/P Values: [50,30.970,2,1,1]**

**dependent=dataset[["charges"]]**

**Output Values:**  **[41748.43]**

**Note: Accuracy No improved from 89%**