***MACHINE LEARNING REGRESSION***

***1.PROBLEM STATEMENT:***

*A client’s requirement is, he wants to predict the insurance charges based on the several parameters. The Client has provided the dataset of the same.*

***SOLUTION:******INSURANCE CHARGES PREDICTION***

***2.BASIC INFO ABOUT THE DATASETS:***

*The dataset contains Age, Sex(Gender),BMI, Smoker, No.of.Children with respect to the insurance charges .*

***Why These Factors Are Important:***

* ***Age****: Health risk increases with age.*
* ***Sex****: Different health risks are associated with men and women.*
* ***BMI****: Reflects health status related to weight.*
* ***Smoker Status****: Smoking drastically increases health risks.*
* ***Number of Children****: Affects the cost of family health coverage.*

***3. PREPROCESSING TECHNIQUES****:*

***1. One Hot Encoding -*** *To convert categorical data into nominal data.*

***2. Standardization -*** *To increase the accuracy level of the input and output dataset which is between (-n to n).*

1. ***MODEL CREATION:***
2. ***Domain Selection :***

*Input is numerical value so that it falls under* ***Machine Learning*** *domain.*

1. ***Learning Selection :***

* *Requirement is clear that it is* ***Insurance charges prediction****.*
* *Both Input & Output columns are present.*

*So, it fall under* ***Supervised Learning****.*

1. ***Regression or Classification Selection:***

*Output Column insurance charges is numerical value so that it falls under* ***Regression.***

*Developed using various Machine Learning Regression algorithms,*

* *Multiple Linear Regression.*
* *Support Vector Machines Regression.*
* *Decision Tree Regression.*
* *Random Forest Regression.*
* *Ada Boost Regression.*
* *LG Boost Regression.*
* ***XG Boost Regression****.*

***5. R2 SCORE FOR EACH MODEL:***

|  |  |  |
| --- | --- | --- |
| ***S.No*** | ***ALOGIRTHM*** | ***R2 SCORE*** |
| *1.* | *Multiple Linear Regression* | *0.79* |
| *2.* | *Support Vector Machines Regression* | *0.8911* |
| *3.* | *Decision Tree Regression* | *0.86* |
| *4.* | *Random Forest Regression* | *0.88649* |
| *5.* | *Ada Boost Regression* | *0.88643* |
| *6.* | *LG Boost Regression* | *0.8951* |
| *7.* | *XG Boost Regression* | ***0.8974*** |

***6. FINAL MODEL****:*

*Algorithm - XG Boost Regression*

*R2 Score - 0.8974*

*Because of the high r2 score I chose XG boost algorithm.*