*Insurance Premium Prediction*

Presentation Subtitle

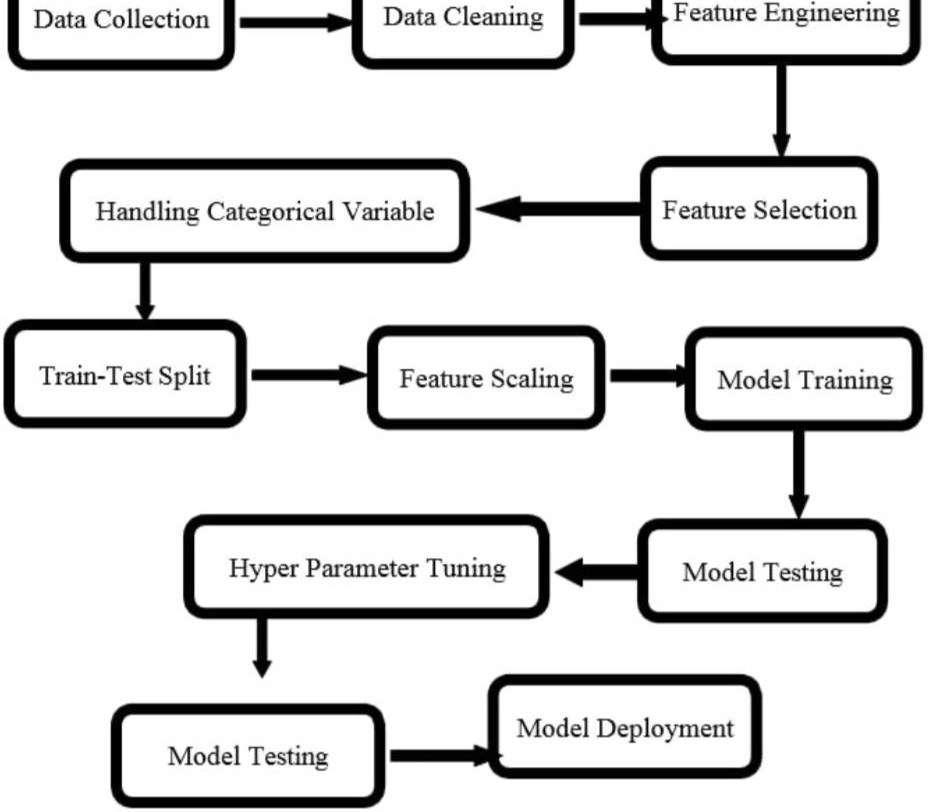
Objective:



Health insurance or medical insurance is a type of insurance that covers the whole or a part of the risk of a person incurring medical expenses. As with other types of insurance is risk among many individuals. By estimating the overall risk of health risk and health system expenses over the risk pool, an insurer can develop a routine finance structure, such as a monthly premium or payroll tax, to provide the money to pay for the health care benefits specified in the insurance agreement. The benefit is administered by a central organization, such as a government agency, private business, or not-for-profit entity. By using machine learning algorithm we are predicting the amount that to be paid by the user.

**Data sharing agreement:**

* Age
* Sex
* Bmi
* No of Children
* Region
* Smoker or not



# Architecture

## Data Preprocessing

1. We cleaned our dataset properly by removing all null value and duplicate value present in dataset.
2. We handled categorical variable by performing One-Hot encoding.
3. We visualized the data using data visualization libraries.
4. We split the data as train and test data.

## Model training

#### We trained our dataset on different Regression Learning algorithm(Linear, Random Forest, Decision Tree Regres sion, XG Boost).

After training the dataset on different algorithms. We got the highest accuracy of 88% on Random Forest Regressor.

Q1) What’s the source of data?

**Q&A**



The data for training is taken from kaggle.

Q 2) What was the type of data?

The data was the combination of numerical and Categorical values.

Q 3) What’s the complete flow you followed in this Project?

Refer slide 5th for better Understanding



**Q&A**

Q 4) What techniques were you using for data pre-processing?

▶ Removing unwanted attributes

▶ Visualizing relation of independent variables with each other and output variables

▶ Checking and changing Distribution of continuous values

▶ Removing outliers

▶ Cleaning data and imputing if null values are present.

▶ Converting categorical data into numeric values.

▶ Scaling the data

Q 5) How training was done or what models were used?

**Q&A**



* As per 80:20 % the training and validation data were divided.
* The scaling was performed over training and validation data.
* Algorithms like Random Forest, XGBoost were used bases on the recall final model was used and we saved that model.



# Thank you