

**Insurance Premium Prediction**

Project Architecture

Domain: Machine Learning

Creator: Vaibhav Joshi

Date: 20.07.2023

# **Architecture**

**Data Preparation**

**Model**

**Development**

**Deployment**

**Deployment**

# **Architecture Description**

## Data Preparation

### Data Description

The health condition data form the insurance company. The goal of this project is to build a prediction model using multiple machine learning techniques and to use a template to document the end-to-end stages. We're trying to predict the expenses the client will make on the premium of insurance.

### Data Preprocessing

In data preprocessing step, we check if there missing data, duplicate values, and data types of each feature. In our dataset, there was not any null and duplicate values

## **Model Development**

### Model implementation

After train and test splitting, pipeline containing Standard Scaler and Ordinal Encoder was fitted to several models such as Linear Regression, Lasso, Ridge, Elastic net, AdaBoost Regressor, Gradient Boosting Regressor, RandomForest Regressor, Their R2 score were obtained. And it was determined that Gradient Boosting performs better than other models.

### Model Evaluation

Test dataset is used to evaluate the model. 30% of dataset was separated for testing. Predicted results of the model are compared with the actual data to check the amount of error.

## **Deployment**

### Designing Form with HTML

For this project, a form is built on HTML.

### Designing a server

A server should be created to run the application continuously. Flask server is built.

### Code deployment on cloud

The codes for this machine learning model should be deployed to the cloud, so that when data is entered into the application, our code runs, and a user gets the result online.

## **Deployment Process**

The code was first committed on Git hub. The pipeline was created between Git and AWS. Then the code was deployed to the AWS.