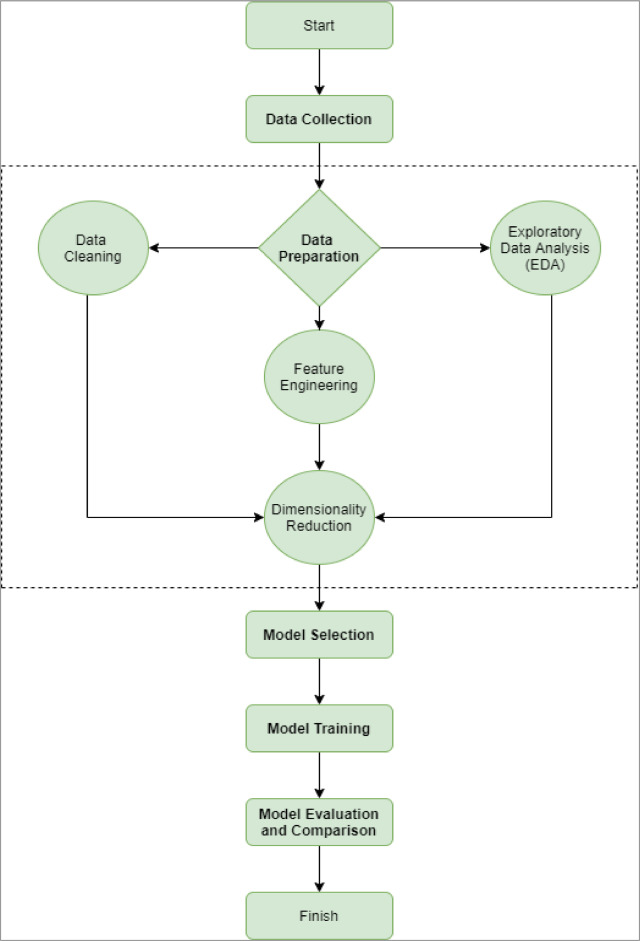
Name:-Saksham Gupta ID No:-E21BCAU0080

Batch: -1 Milestone 3

• Flowchart/ Arch Diagrams: -

x

• **Flowchart explanation**

START -> Gather Medical Insurance Data

Gather all the relevant medical insurance data including the individual's age, gender, BMI, smoking status, number of dependents, region, and charges.

-> Perform Data Preprocessing

Clean and preprocess the gathered data. This includes checking for missing values, removing outliers, normalizing or scaling the data, and encoding categorical variables.

-> Split Data into Training and Testing Sets

Divide the preprocessed data into training and testing sets. The training set will be used to train the machine learning model, and the testing set will be used to evaluate the model's accuracy.

-> Choose Machine Learning Algorithm

Select the appropriate machine learning algorithm for the problem. This could be a regression algorithm such as linear regression, polynomial regression, or decision tree regression.

-> Train the Model

Train the selected machine learning algorithm using the training set.

-> Evaluate the Model

Evaluate the trained model's performance using the testing set. This could involve calculating metrics such as mean squared error, root mean squared error, R-squared, or mean absolute error.

-> Optimize the Model

If the model's performance is not satisfactory, adjust the model's hyperparameters to improve its performance. Hyperparameters could include the learning rate, regularization parameters, number of hidden layers, or number of neurons in each layer.

-> Deploy the Model

Once the model's performance is satisfactory, deploy it for use in making medical insurance cost predictions.

END

**• Algorithm/technique/model related explanation.**

Linear regression is a type of machine learning algorithm that is often used for predicting numerical values. In the context of medical insurance cost prediction, linear regression can be used to predict the cost of an individual's medical insurance policy based on various factors such as age, gender, BMI, smoking status, number of dependents, and region.

To use linear regression in your project, you would first need to gather and preprocess the relevant medical insurance data. This would involve cleaning the data, removing any outliers, and encoding any categorical variables.

Next, you would split the preprocessed data into training and testing sets. The training set would be used to train the linear regression model, while the testing set would be used to evaluate the model's performance.

To train the linear regression model, you would use the training data to find the coefficients for the linear equation that best fits the data. This equation would take the form of:

y = b0 + b1x1 + b2x2 + ... + bnxn

where y is the predicted medical insurance cost, x1, x2, ..., xn are the input variables (such as age, gender, BMI, smoking status, number of dependents, and region), and b0, b1, b2, ..., bn are the coefficients that the model learns during training.

Once the model is trained, you would use the testing data to evaluate the model's performance. This would involve calculating metrics such as mean squared error (MSE), root mean squared error (RMSE), or R-squared.

If the model's performance is not satisfactory, you may need to adjust the model's hyperparameters, such as the learning rate or regularization parameters, to improve its performance. Once the model's performance is satisfactory, you can deploy it for use in making medical insurance cost predictions.

In summary, linear regression can be used in your Medical Insurance Cost Prediction project to predict the cost of an individual's medical insurance policy based on various factors. The algorithm works by finding the coefficients for a linear equation that best fits the data, and then using this equation to make predictions.

• **Dataset description**

