Flow of your approach -

**1]Understanding -Problem statement**

Ans : Here we have to identify that how accurately is ML model is predicting that given transaction is fraudelent trandsaction or not.

**2]Perform exploratory data analysis**

**Ans:** As a part of exploratory data analysis, we have performed followings steps:

a] Identifying null values : As observed there is no null values are available in credit card fraud detection dataset.

**3]Preprocess the data:**

**Ans:** a] Identified dataset is imbalance dataset or not : There are 91 % transaction are non fraudulent transaction and 8% transaction are fraudulent transaction so this is not a imbalance dataset.

b]Feature selection : We have selected all 8 feature for this problem.

**4]Select Training data, test data:**

**Ans:** We have selected 80% of test data and 20% of train data for our problem statement.

**5]Train the model**

**Ans:** We have trained model for following algorithm:

a]Linear regression

b]Logistic regression

**6]Test the model (Predictions and reporting):**

&

**7]Evaluate the model performance**

**Ans:**

a]For Linear Regression :Model has accuracy of 92%.

Confusion Matix :

[[160430 390]

[ 13338 969]]

Report :

Precision = 0.713

Recall = 0.068

Accuracy = 0.922

F1 Score = 0.124

b] For Logistic Regression: Model has accuracy of 96.7%.

Confusion Matix :

[[160763 57]

[ 5670 8637]]

Report :

Precision = 0.99

Recall = 0.965

Accuracy = 0.967

F1 Score = 0.97

**8]Suggest ways of improving the model** : If user used random forest technique then model accuracy can be increased. As we change number of trees in random forest technique then accuracy of the model also will get changed.

**Que: State all your assumptions clearly**

**Ans :** 1] For this problem I am considering all features to predict result.

2]Given dataset is not a imbalnce dataset

3]I am assuming Logistic regerssion will be best model to get result.

**Que :** **Provide clear explanations to justify your stand**

1] As this data set is not a imbalance dataset hence accuray will get more important than precision value.

2] Result of the given dataset is a binary value which means this is a binary classifier and for binary classifier logistic regersssion is the best model.