**Write Up**

1. **Import Libraries and read dataset**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

train\_data = pd.read\_csv("ML\_train.csv")

test\_data = pd.read\_csv("ML\_test.csv")

1. **Task-1 if for any column(s), the variance is equal to zero, then you need to remove those variable(s).**

Step 1: Find Variance = 0

Step 2: Than list Out

Step 3: Drop it

1. **Task-2 Check for null and unique values for test and train sets.**

Step 1: Find null value and it found False

Step 2: Than find no. of Unique value of train\_data

Step 3: Than find no. of Unique value of test\_data

1. **Task-3 Apply label encoder.**

Step 1: List out non-numerical column

Step 2: List out the unknown class

Step 3: Label it unique unknown class

1. **Task-4 Perform dimensionality reduction.**

Step 1: Drop unwanted components

Step 2: Use PCA Method for variance

Step 3: By PCA Variance plot Gaph

Step 4: Check the useful Component

1. **Task-5 Predict your test\_df values using XGBoost.**

Step 1: Impoet XGBoost library

Step 2: Split the train and test data.

Step 3: Find r2\_score

Step 4: Find RSME of train and test data.

Step 5 : Plot the graph Actual vs Predicted on Train data