Report of DS-Intern-Assignment--Sudeep\_Sapkota for Mechademy

1. In data preprocessing step, I first removed the timestep data , but later found its importance to improve R^2 to 0.091 for random forest model
2. The R^2 value is quit less and model is performing very poor

I tried to remove the outlier but it resulted in quite a loss in the data points **( 16k datapoints to 1k datapoints !!!)**

I got R^2 =0.89 by removing outlier by using IQR method.

1. Very less correlation is obtained between the features and the target variable leading to underfit.

At last Thanks for this opportunity to solve this assignment as it improved my hand-on on ML workflow.

Sudeep Sapkota

MTech final year student,

Mechanical Engineering Department

IIT Bombay