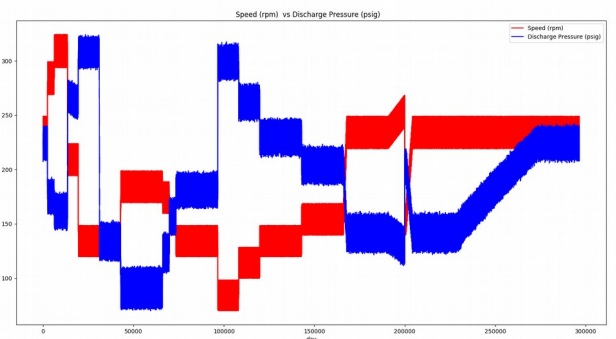
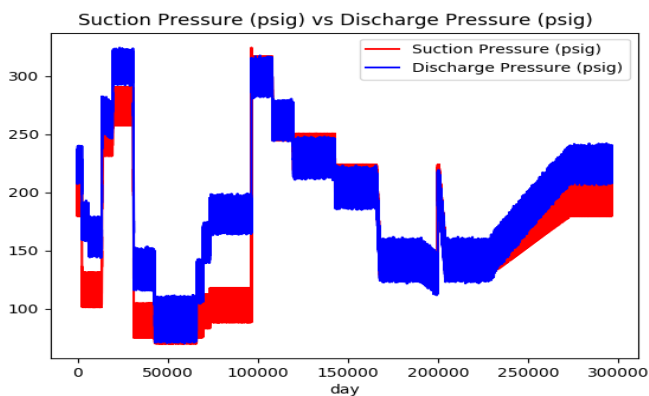
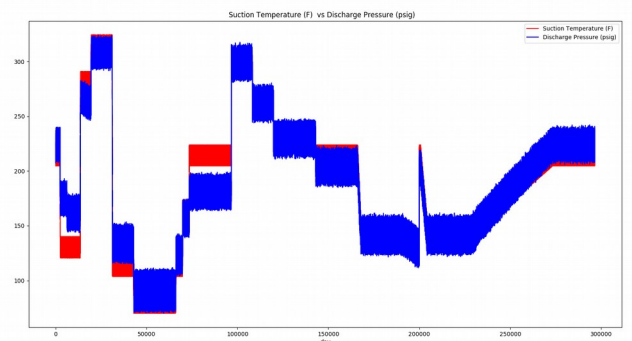
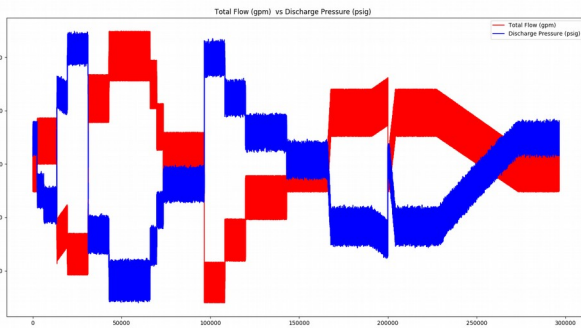
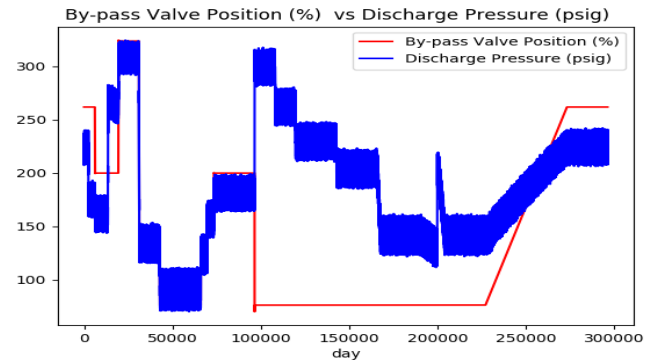
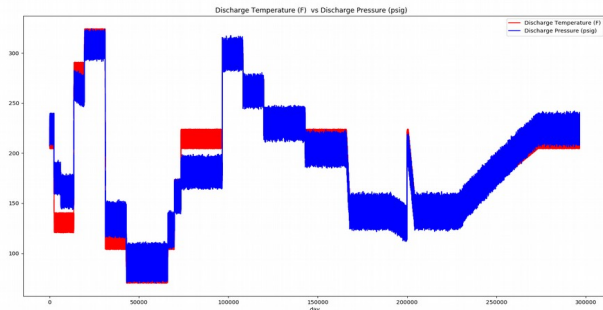


DATA PREPROCESSING

1. The only data preprocessing I did was to scale all the columns to the same range as the discharge pressure.

DATA ANALYSIS

The plot for each feature vs discharge pressure is plotted and checked for redundant features.



As it can be seen, all the features, either directly or inversely effects the discharge pressure. Hence no feature selection is required.

ALGORITHM

I have used 3 algorithms for this problem.

1. Linear Regression. 2. Random forest. 3. ANN

	Complexity of algo	Time Taken	Error on test set
Linear Regression	Very simple	Very Low	6.03
Random Forest	Simple	Very Low	34.74
ANN	Complex	High	5.08

CONCLUSION

1. Given that discharge pressure is correlated to most of the features linearly, Linear regression works fine for the problem. Moreover, it takes extremely less time and hence is a good choice of algorithm. As Linear regression gives good result on the error and takes negligible time, it is the algorithm that should be opted in case where the degree of accuracy can be compromised but time is of essence.

2. However, ANN gives better accuracy. The network that I used has a single hidden layer with 6 nodes. The batch size considered is 10 and number of epoch is 200. Adam is used as the optimizer. Making the network more complex, or changing the hyper parameters can yeild a much better error. So, in situations where accuracy is of importance, ANN can be chosen.