INTRODUCTION

Marine gas turbines are used in naval propulsion plants and, as such, are required to be extremely reliable. In this application, condition-based maintenance is highly preferable to corrective maintenance, due to its lower costs and shorter down times.

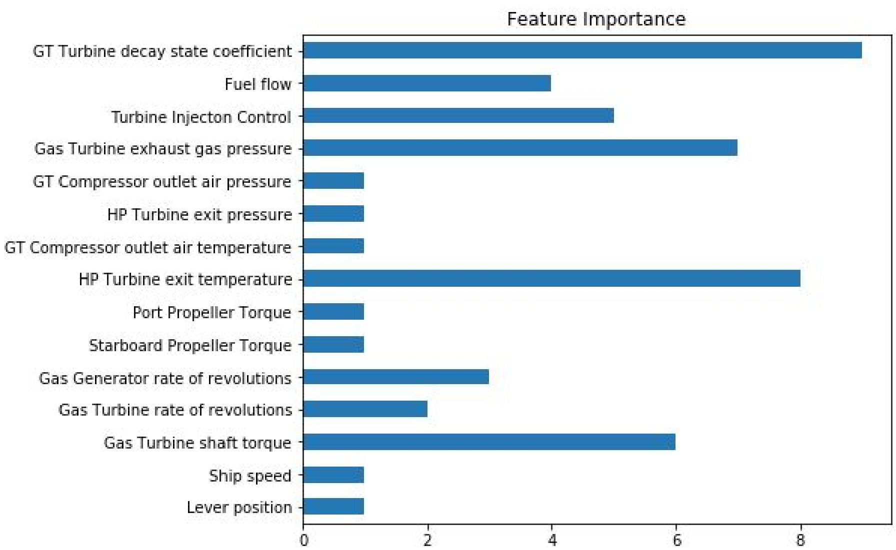
This report summarizes the findings from modeling gas turbine compressor deterioration using a dataset collected in a simulator of a Combined Diesel Electric and Gas propulsion plant mounted in a frigate.

Area: Condition-Based Maintenance in Naval Domain
Format Type: Each feature vector is a row on the text file (16 elements in each row)
Does your data set contain missing values? No
Number of Instances (records in your data set): 11984
Number of Attributes (fields within each record): 16

OBSERVATIONS

Dataset- The data is highly correlated as parameters of Gas Turbine are interconnected.

With high correlation, we expect multicollinearity problems .The ordinary least square regression produces a model with high collinearity. Feature selection method is used to reduce high collinearity -



Further lasso regression or l1 regularization is implemented to get most significant features.

PREDICTIVE MODELS

We can avoid multicollinearity as the primary purpose of our regression analysis is to predict a new response. Linear regression,Support Vector Regression and Random Forest models are evaluated and compared based on cross validation accuracy and root mean squared error metric.

Plotting Cross Validated Predictions

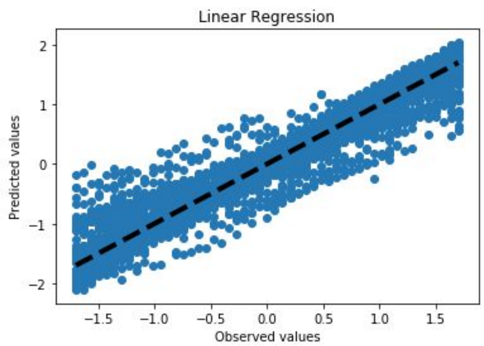
Below figures visualize prediction errors for each model and shows which one is a better estimate.

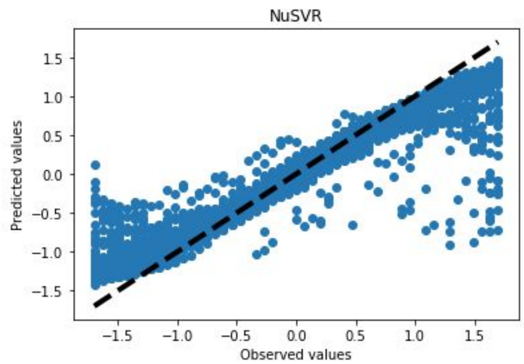
The input dataset is split into 70% of training and 30% testing data i.e, 8353 points of traingd data and 3581 points of testing data of all the models.

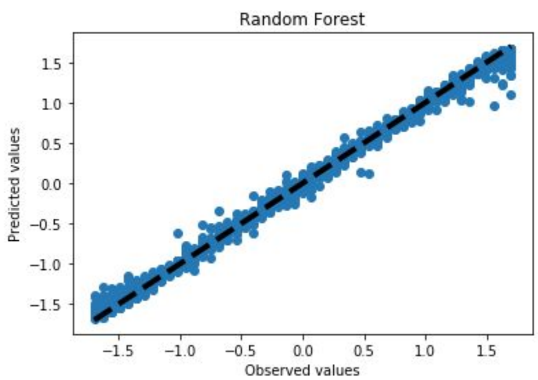
Linear Regression - All the features of the training data were used to fit Linear regression and test data is used to predict the deterioration and the Linear Regression plot below compares observed values and predicted values from the test data. The accuracy score(R-squared) of this model gave 88.7% accuracy which is good. But from the cross validated prediction plots between the actual values and the predicted values(predicted from the test data) show high variance which makes the model a bad fit to predict the deterioration.

SVR - Initially, the effects of Kernel term variations on SVR model performance were evaluated.Exhaustive search is made for regularization term for an estimator. With C=0.01 and Kernel ='linear' SVR is modeled with cross validation score(R-squared) of 99.7%. The cross validated prediction plots between the actual values and predicted values from test data show that SVR model works poorly with high and low values and maybe we need to try wider range of parameters.

Random Forest - Random forest gives much more accurate predictions when compared to SVR and Linear Regression.It takes preprossed data and hyperparameters tuned so that we don't observe any high variance from the model. The model fits the data perfectly and give a accuracy score(R-squared) 0f 99.7%. The cross validated prediction plots shows that the actual values and the predicted values have good correlation with not many outliers and low variance.







Model	Cross validation Prediction Accuracy	Root Mean Squared Error (RMSE between observed and predicted values)
Linear Regression	89.9%	0.316
NuSVR	88.4%	0.339
Random Forest	99.7%	0.05

CONCLUSION

The results showed that it is possible to predict when the maintenance needs to be planned based the values given by combining features and also most significant features to be monitored - HP Turbine exit pressure, GT Compressor outlet air pressure ,Gas Turbine exhaust gas pressure, Fuel flow. We also see that the Random Forest prediction based on all variables provide a benchmark of low the prediction error.