Application of Machine Learning Technique to Predict Crude Distillation Column Inlet Temperature / Furnace Coil Outlet Temperature in Order to Maximize Distillate Yield and to Minimize Fuel Firing in Furnaces

Debdatta Kundu
K. J. Somaiya College of Engineering, Mumbai, India

Tejas Khanolkar
K. J. Somaiya College of Engineering, Mumbai, India

Tirth Shah
K. J. Somaiya College of Engineering, Mumbai, India

Sarvesh Bangad
K. J. Somaiya College of Engineering, Mumbai, India

ABSTRACT

The optimization of furnace firing in a Crude Distillation Unit (CDU) helps refineries to save fuel and to boost up refinery margin by increasing distillate yield. The paper is focused on development of a suitable model to predict Crude Distillation Column inlet temperature / furnace coil outlet temperature (COT) of a petroleum refinery by using Machine Learning techniques. Different regression algorithms are used to fit the given data and error functions are computed for the different models. Their performance is then compared to select the best performing model. The models are developed based on actual operating data from the Crude Distillation Unit of an existing petrochemical refinery and the outputs are tested to predict the optimum range of COT and Random Forest Regressor is found to be the best model for predicting optimum Furnace COT values of a Crude Distillation Column based on the given features.

General Terms

Machine Learning, Crude Distillation Column.

Keywords

Machine Learning, Data Processing, Regression Model, Crude Distillation Column, Furnace Coil Outlet Temperature, Refinery Margin.

1.   INTRODUCTION

Refinery processes various types of crude oils, some crudes are light and some crudes are heavy (low API). Some crude have high sulfur (Sour crude) content and some crudes have low sulfur (Sweet crude) content. Total Acid Number (TAN) number, Asphaltene content and Nitrogen content of different types of crude oil are also different. Therefore there is wide variance in properties of different types of crude oil depending upon sources. Based on the crude oil properties different types of crude oil give different types of yield.[1] In any refinery Crude oil first processes in a Crude Distillation Column where crude is being heated up in a furnace before separating the crude oil into different fractions inside the distillation column.

Crude Distillation Unit (CDU) is the first unit of any refinery. CDU normally consists of two distillation columns – Crude Distillation Column followed by a Vacuum Distillation Column. The different types of products obtained from a Crude Distillation Unit are:

1. Fuel Gas (FG)
2. Liquefied Petroleum gas (LPG)
3. Naphtha (Light and Heavy Naphtha)
4. Superior Kerosene (SK) and Aviation Turbine Fuel (ATF)
5. High Speed Diesel (HSD)
6. Waxy
7. Fuel Oil

The objective of the project is to predict Crude Distillation Column inlet temperature (furnace coil outlet temperature) by using Machine Learning technique to maximize distillate yield and to minimize fuel firing in furnaces.

A total of nine important parameters of Crude Distillation Column shown in Figure 1 and its associated equipment were chosen for modelling purpose by using Machine Learning techniques so as to predict the optimum furnace coil outlet temperature (COT). By predicting optimum COT, furnace firing can be optimized and distillate yield can be maximized to boost up refinery margin.

Machine Learning is an application of Artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine Learning focuses on the development of computer programs that can access data and use it to learn for themselves[2]. The Machine Learning models were tested on a data set in an existing refinery in India and the results were found very much enriching.

Figure 1 represents the typical schematic diagram of a Crude Distillation   Column
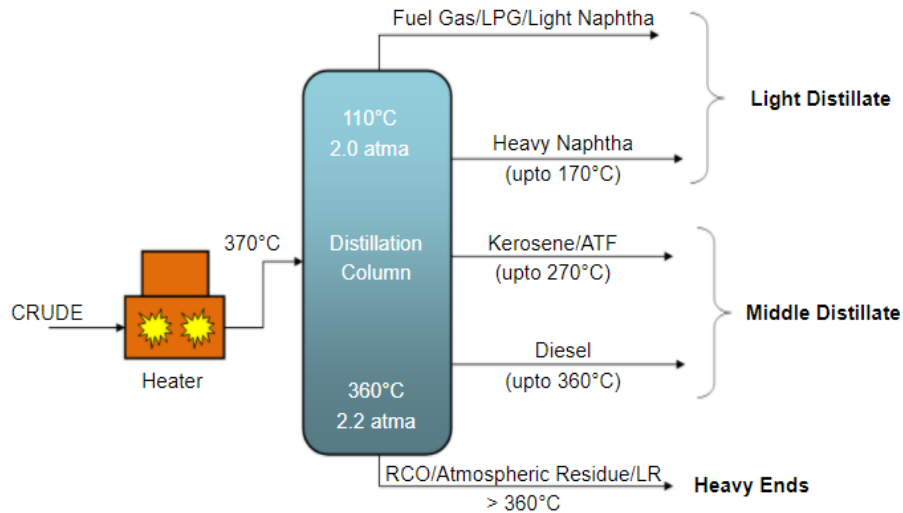
Figure 1 Crude Distillation Column

Following are the major important parameters considered to build the Machine Learning models

1. Feed
2. Reflux Temperature
3. Density
4. Over flash Flow
5. PHT1 Temperature
6. Desalter Temperature
7. Preflash Inlet Temperature
8. 3rd PHT Inlet Temperature
9. Final PHT (Preheat train) Temperature
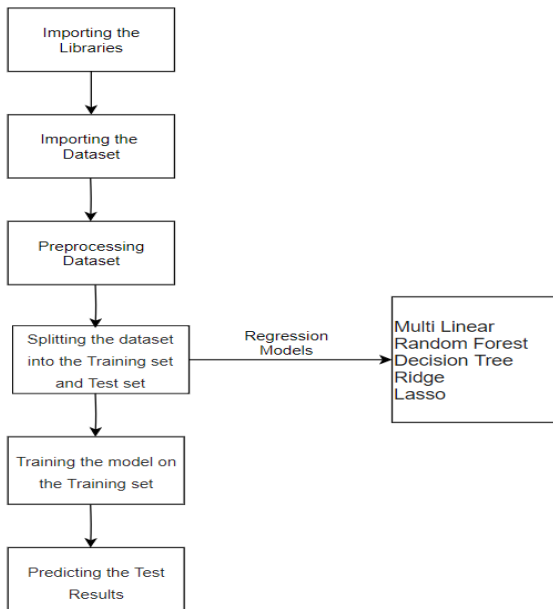
2. METHODOLOGY

Figure 2 Flowchart for Machine Learning Process

Figure 2 shows the flowchart for an overview machine learning process consisting of all the steps required for applying machine learning algorithms for the refinery process. First step of the process is importing all the necessary libraries and the dataset. Pre-processing is applied on the dataset and it is split into training and testing sets. Different Machine Learning algorithms are applied on the dataset and the results are evaluated.

There are many models that researchers and data scientists have created over the years. Dataset can be trained in many ways, by using different models of regressions. Five regression models have been used for this project: -

- Multiple Linear Regression
- Random Forest Regression
- Decision Tree Regression
- Ridge Regression
- Lasso Regression

3. RESULTS AND DISCUSSIONS
Results and comparison of the regression models on evaluation metrics.

3.1 Multiple Linear Regression
The multiple linear regression model is the most commonly applied statistical technique for relating a set of two or more variables.[3]
The evaluation metrics obtained are as follows:

Table 1 Evaluation metric for multiple linear regression

Evaluation Metric	Value
Root Mean Squared Error	2.6546
R2 Score	0.6836

R-Squared(R2 Score) is a statistical measure of how close the data are to the fitted regression line while Root Mean Squared Error(RMSE) is the square root of the mean of the squared error. Closer is the value of R2 score to 1 and lower the value of RMSE better is the model's performance. Table 1 shows the RMSE and R2 score for Multiple Linear Regression. The RMSE value obtained is relatively high and the R2 score is low indicating an average performance of the model on the testing dataset.
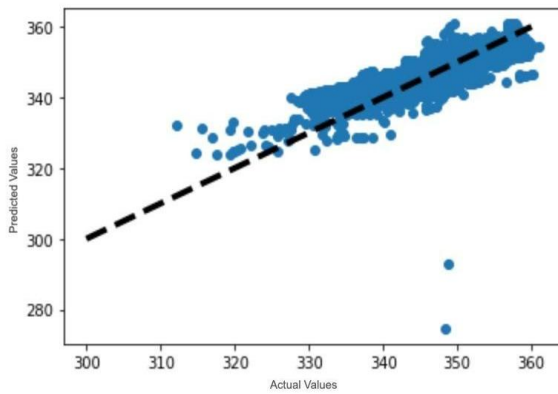
Figure 3 Actual Values v/s Predicted Values for Multiple Linear Regression

Figure 3 Shows the distribution of actual values and predicted values for Multiple Linear Regression and it can clearly be seen from the figure that the predicted values are scattered and deviate from the best-fit line indicating an average performance of the model.

3.2 Random Forest Regression

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.[4] Random decision forests correct for decision trees' habit of over fitting to their training set.

Table 2 Evaluation metric for random forest regression

Evaluation Metric	Value
Root Mean Squared Error(RMSE)	0.3721
R2 Score	0.9937

Table 2 shows the RMSE and R2 score for Random Forest Regression. The RMSE value obtained is very less and the R2 score is very close to 1 indicating that the model performed extremely well on the testing dataset.
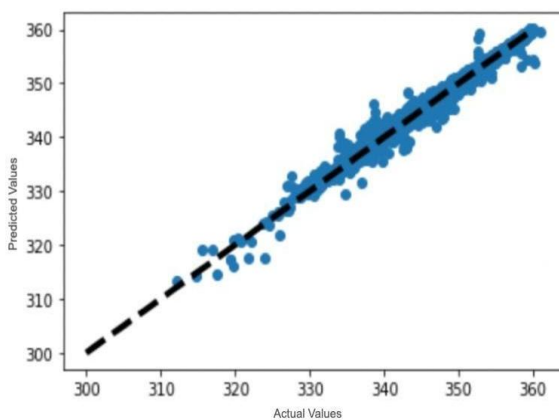

Figure 4 Actual Values v/s Predicted Values for Random Forest Regression

Figure 4 Shows the distribution of actual values and predicted values for Random Forest Regression and it can clearly be

seen from the figure that the predicted values are nearer to the best-fit line indicating an excellent performance of the model.

3.3 Decision Tree Regression

Decision tree as the name suggests is a tree structured predictive model. They are one of the most powerful tools in data mining and machine learning, as it is very easy for a person to derive conclusions from the model's result. A decision tree is one of the popular and easily understood algorithms as its results mimic the human brain, hence are easy to understand and derive rules out of these results.[5]

Table 3 Evaluation metric for decision tree regression

Evaluation Metric	Value
Root Mean Squared Error(RMSE)	0.5953
R2 Score	0.9840

Table 3 shows the RMSE and R2 score for Decision Tree Regression. The RMSE value obtained is very less and the R2 score is very close to 1 indicating that the model performed well on the testing dataset.
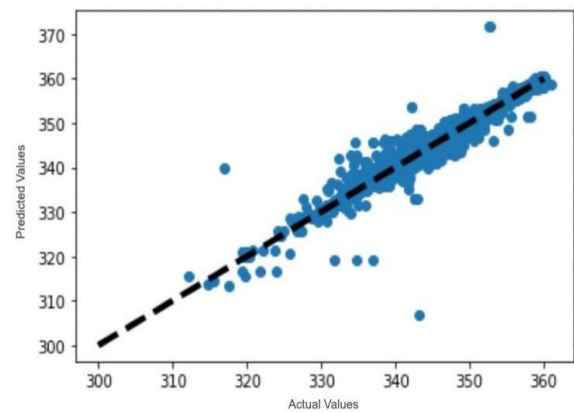

Figure 5 Actual Values v/s Predicted Values for Decision Tree Regression

Figure 5 shows the distribution of actual values and predicted values for Decision Tree Regression and it can clearly be seen from the figure that the predicted values are nearer to the best-fit line indicating an excellent performance of the model.

3.4 Ridge Regression

Ridge regression is a method for analyzing the data which are affected by multicollinearity. If multicollinearity is present in the data, least squares estimates are unbiased, but the variances are large hence they are far away from the true value. Ridge regression is identical to least squares, unless the ridge coefficients are estimated by minimizing a slightly different quantity. It is hoped that the net effect will be to give estimates that are more reliable.[6]

Table 4 Evaluation metric for ridge regression

Evaluation Metric	Value
Root Mean Squared Error(RMSE)	2.6546
R2 Score	0.6836

Table 4 shows the RMSE and R2 score for Multiple Linear Regression. The RMSE value obtained is relatively high and the R2 score is low indicating an average performance of the model on the testing dataset.
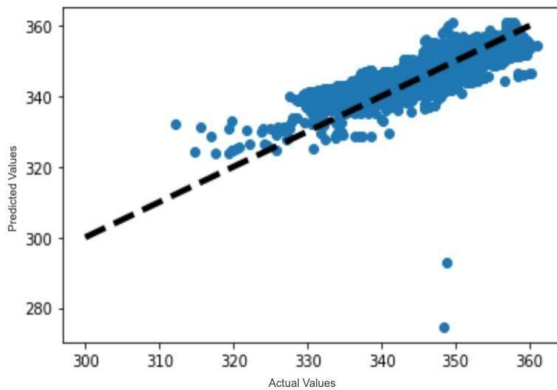

Figure 6 Actual Values v/s Predicted Values for Ridge Regression

Figure 6 shows the distribution of actual values and predicted values for Ridge Regression and it can clearly be seen from the figure that the predicted values are scattered and deviate from the best-fit line indicating an average performance of the model.

3.5 Lasso Regression

Lasso regression is a type of linear regression that uses shrinkage. Shrinkage is where data values are shrunk towards a central point, like the mean. This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination. The acronym "LASSO" stands for Least Absolute Shrinkage and Selection Operator.

LASSO improves both prediction accuracy and model interpretability by combining the good qualities of ridge regression and subset selection.[6]

Table 5 Evaluation metric for lasso regression

Evaluation Metric	Value
Root Mean Squared Error(RMSE)	2.6546
R2 Score	0.6836

Table 5 shows the RMSE and R2 score for Multiple Linear Regression. The RMSE value obtained is relatively high and the R2 score is low indicating an average performance of the model on the testing dataset.
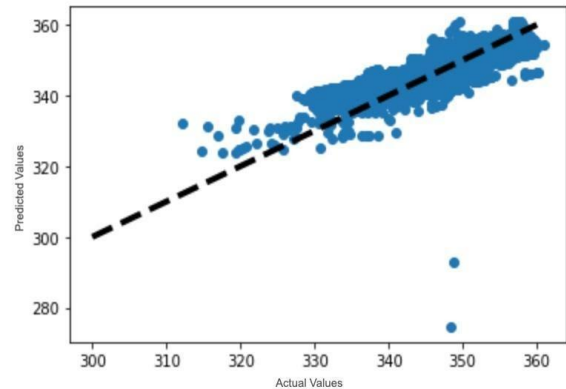

Figure 7 Actual Values v/s Predicted Values for Lasso Regression

Figure 7 shows the distribution of actual values and predicted values for Lasso Regression and it can clearly be seen from the figure that the predicted values are scattered and deviate from the best-fit line indicating an average performance of the model.

4. CONCLUSION AND FUTURE SCOPE

Table 6 Performance analysis of various regression models

Regression Models	Root Mean Square Error (RMSE)	R2 Score
Multiple Linear Regression	2.6546	0.6836
Random Forest Regression	0.3721	0.9937
Decision Tree Regression	0.5953	0.9840
Ridge Regression	2.6546	0.6836
Lasso Regression	2.6546	0.6836
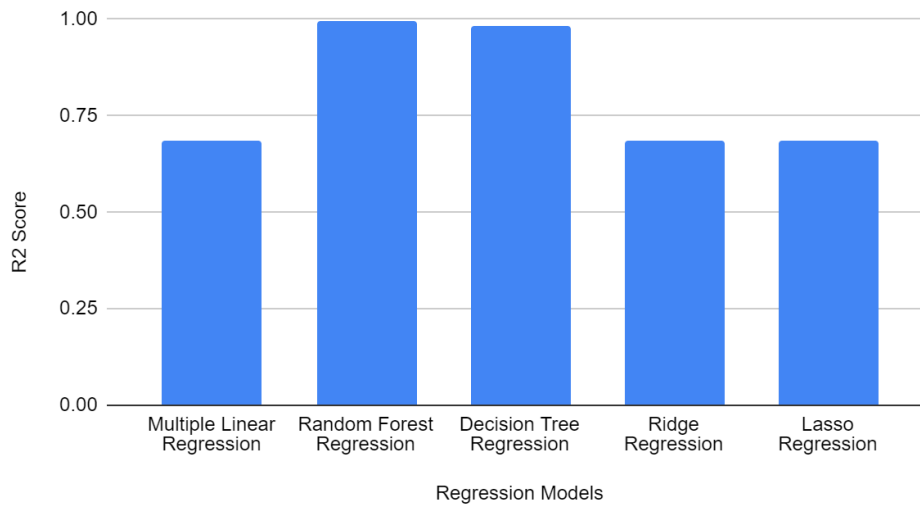
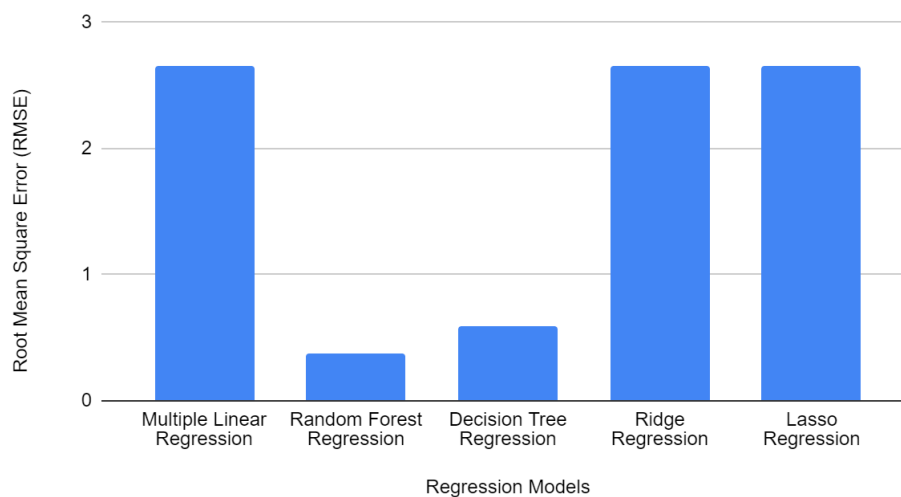**Figure 8 R2 score of various regression models using graph**

Figure 9 RMSE score of various regression models using graph

The aim of this project was to predict optimum FURNACE COT based on certain features. The data was trained on 5 regression models namely Multiple Linear Regression, Random Forest Regression, Decision Tree Regression, Ridge Regression, Lasso Regression. After evaluating all five models on the dataset, their performances were compared to find out which model predicts better RMSE and R2 Scores are tabulated for all the models.

The R2 score and RMSE (Root mean squared error) of the respective models were calculated based on the models and it was observed that the Random Forest Regressor fits extremely well on the data with an R2 score of **0.9937** and RMSE value of **0.3721.** (See figure 8, 9 and table 6)

Thus Random Forest Regressor was found to be the best model for predicting optimum FURNACE COT values of a Crude Distillation Column based on the given features.

The results obtained from the Random forest regression model are very satisfying but there are a lot of other models which could be tested on the given dataset. Deep Learning Algorithms could also be applied and may lead to different results. There might be other industrial processes or techniques which could be better and other models or algorithms may tend to work better on them. Future work concerns deeper analysis of particular mechanisms, new proposals to try different methods, or simply curiosity.

5. REFERENCES

[1] Nicholas P. Cheremisinoff, Paul Rosenfeld ,"Chapter 1 - The petroleum industry", Handbook of Pollution Prevention and Cleaner Production - Best Practices in The Petroleum Industry, William Andrew Publishing,2009,Pages 1-97,ISBN 9780815520351,https://doi.org/10.1016/B978-0-8155-2035-1.10001-6.
(http://www.sciencedirect.com/science/article/pii/B9780815520351100016)

[2] S. Ray, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 35-39, doi: 10.1109/COMITCon.2019.8862451.

[3] Jobson J.D. (1991) Multiple Linear Regression. In: Applied Multivariate Data Analysis. Springer Texts in Statistics. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-0955-3_4

[4] Ho, Tin Kam (1995). *Random Decision Forests* (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282. Archived from the original (PDF) on 17 April 2016. Retrieved 5 June 2016.

[5] S. Pathak, I. Mishra and A. Swetapadma, "An Assessment of Decision Tree based Classification and Regression Algorithms," 2018 3rd International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2018, pp. 92-95, doi: 10.1109/ICICT43934.2018.9034296.

[6] R. Muthukrishnan and R. Rohini, "LASSO: A feature selection technique in predictive modeling for machine learning," *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, Coimbatore, 2016, pp. 18-20, doi: 10.1109/ICACA.2016.788791