International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 09 | Sep 2022 www.irjet.net p-ISSN: 2395-0072

Machine Learning Based Prediction Model to Estimate Calorific Value of Diesel

Ankur Harge[1], Devendra Aradhye[1], Shirish Chougule[1], Sanveg Jain[1], Milankumar Nandgaonkar[2]

[1]Mechanical Engineering, College of Engineering Pune, 411005 India
[2]Associate Professor, Dept. of Mechanical Engineering, College of Engineering Pune, 411005 India

---

Abstract - *Petroleum as obtained from different refineries, is predominantly a mixture of many hydrocarbons with differing molecular structure. This mixture varies from company to company, refinery to refinery due to difference in there blending techniques. It may also contain small amounts of Sulphur, oxygen, nitrogen and impurities such as water. This influences the chemical and physical properties of our fuel. And one such important property is Calorific value. Calorific value of fuel is an important property thereby measuring the energy content as well as purity of fuel. The conventional instrument used for measuring calorific value of fuel is bomb calorimeter. Measuring calorific value by bomb calorimeter is a destructive technique where fuel samples are burned every time, we measure the value. In order to overcome this disadvantages, alternate way of determining calorific value of liquid fuels is elaborated here. Prediction model based on different Machine Learning (ML) techniques which will successfully predict the calorific value of diesel are developed. The dataset originating from experimental measurements consisting of different properties of diesel were gathered, processed and then used as an input for machine learning model. Artificial neural network (ANN), Support vector regression (SVR) and multivariate linear regression (MLR) are the various regression models used for predicting calorific value of fuel followed by comparative study between these techniques on the basis of accuracy, robustness and reliability.*

Key Words: **Root Mean Square Error, Machine Learning, Artificial neural networks, Support Vector Regression, Multivariate Regression**

1.INTRODUCTION

Calorific value (CV) is an important measure to check performance and purity of fuel. Existing methods to determine calorific value given in IS 1448 [P-6] and [P-7] takes more time and are expensive. And sometime the failure of instruments as demonstrated by John Nail, et al [1] can be risky. Plus, it is economically impossible to assert the calorific value of various fuel samples originating from different refineries. Hence many of the researchers were interested in determining the calorific value of fuel using indirect methods like establishing correlation between calorific and other parameters – chemical, physical or spectral.

W. F. Faragher et al in 1970 had developed a quantitative relationship between the calorific value and the A. P. I. gravity of fuel oils such as topped crude oil and cracked residuum [2]. Oliveira L. E. and Da Silva M. work aimed to obtain experimentally and compare the cetane number and calorific value of different blends of biodiesel [3]. While Rajneesh K. Verma, et al used refractive index to detect the adulteration of fuel by SPR based fiber optic technique [4]. Feasibility study was carried out for measuring gross calorific value (GCV), carbon content, volatile matter content and ash content of solid biomass fuel using laser-induced breakdown spectroscopy by Zhimin Lu et al [5]. PLS model based on spectra and Z-score standardization established an equation for prediction of calorific value of biomass. Also, numerous efforts on quantitative analysis have been done using NIR and the multivariate method of partial least squares (PLS) regression [6-9], especially for diesel fuels - fraction of petroleum atmospheric distillation 200–300 °C [10]. Several research analyzed biodiesel/diesel blends based on NIR spectroscopy and diverse multivariate methods have produced interesting results [11–14]. Particularly, Alves et al. demonstrated, the simplicity of using PLS versus others multivariate methods, as e.g., Support Vector Machine (SVM) [14]. Nevertheless, an important challenge that still has to be faced is the cost and complexity involved in these techniques. Achieving such performance depends on diverse chemo metrical calibration settings. Some of them, as prediction intervals, infrared spectral region, calibration algorithm, plus, spectral pre-processing procedure have to be studied for petrochemicals [15-17]. This encouraged us to go for a easier approach and check for correlation between calorific value and other properties like density, viscosity, pH, refractive index, etc.

Also, prediction model which will correlate these properties to the calorific value was to be developed. There were many researchers who had successfully predicted calorific value using machine learning models. Like Shagufta U. Patel et al in 2006 had developed total of seven nonlinear models using the ANN methodology for the estimation of GCV with a special focus on Indian coals [18]. The most accurate model had RMSE value of 0.514, Average percentage error 2.067 and coefficient of correlation 0.997 for test data set.

In 2012 Colm D. Everard et al developed a PLSR model to determine calorific value of dedicated bioenergy crops [19].

Visible and near infrared spectroscopy was used to determine the composition of 44 samples. While Qihong Feng et al [20] carried a comparative study between 3 regression model to determine which gives highest accuracy while predicting calorific value of coal 1) support vector machine 2) alternating conditional expectation and 3) artificial neural network are the 3 models taken into consideration. It was found that SVM and ANN regression model can work on small dataset and gives results with high accuracy. In [21] Lijun Xu et al also carried similar study but used single technique with three different approaches. Several statistical approaches including principal component analysis (PCA), independent component analysis (ICA) and partial least squares analysis (PLSA) were used in SVM nonlinear regression analyses. Work by S.S. Matin and S. Chehreh Chelgani showed that random forest method gives better results only for large and multidimensional database [22].

Researchers used machine learning models to predict physical properties of fuels like coal, biodiesel and biomass. So, we have extended the methodology and applied it to diesel and facilitated prediction of it's CV. Also, from these literatures it was found that the calorific value is best predicted by non-linear regression models and hence depending upon various factors like size of data, complexity, performance, ANN, SVM, MVR techniques were selected. These techniques gave most accurate results while predicting calorific value of diesel. Thus, a relatively less complex technique of calorific value measurement of diesel has been proposed.

1.1 Data collection

All samples were of post-blending, ready to use diesel, which were collected from the commercial outlets of different petroleum companies in Maharashtra, India like HPCL, IOCL, BPCL, Reliance, Essar. The samples composition included the addition of diverse amounts of kerosene, atmospheric and heavy atmospheric diesels from various refining processes such as fluid catalytic cracking (FCC), direct distillation, coking, Vis breaking and hydrotreatment. Their greatest source of variability was their composition, which was essentially dependent on the origin of the crudes. Usually, hydrocarbon streams employed to formulate the diesel fuel can be derived from 2 to 5 crudes from diverse sources as well as streams from various refining processes. Outlets in Maharashtra receive diesel out of refineries from Gujarat, Bina, Mumbai, Vishakhapatnam, Chennai, etc. So, to have equal inclusion of all refineries and outlet brands and to avoid development of biased model, extensive number of samples were collected. This took in to account all the variations possible. A total of 179 samples were used to construct, validate and assess the models. Sample each of size 400 ml approx. were collected. Government of India's guidelines for storing of diesel (class B fuel) as per "The Petroleum Rules 1976" were followed. Feature extraction followed by pre-processing the data was carried out over each sample.

1.2 Feature extraction and selection

The basis for selecting a feature was ease with which a feature can be measured, possibility of correlation with CV, measurement accuracy of the instrument or technique and previous studies. Degree of correlation could be confirmed only after pre-processing the collected data. So, other factors were taken into account while selecting a feature. Properties that were selected were: pH, Viscosity(cSt), Density(kg/m^3), Refractive Index, Flash Point(˚C) , Fire Point(˚C). Though this list is not at all exhaustive, but including many features can increase complexity of the model, whose results would be difficult to explain or correlate. Problem that can arise when too many features are considered is that of multicollinearity. And further methods like PCA or Factor Analysis need to be employed to facilitate dimension or redundancy reduction. Though above tests were still accounted on the selected features to eliminate any uncertainty.

Table -1: Precision of Reference methods

Parameter	Method	Repeatability (r)	Reproducibility (R)
Calorific Value (KJ/kg)	IS 1448 [P: 6]	276 J/g	773 J/g
pH	EN ISO 10523	-	-
Viscosity(cSt),	IS 1448 [P: 25]	0.0035	0.007
Density (g/ml^{-1})	EN ISO 12185	0.0002	0.0005
Refractive Index	EN ISO 22241	0.004	0.01
Flash Point	EN ISO 2719	1.62–2.03	3.98–4.97
Fire Point	EN ISO 2719	1.62–2.03	3.98–4.97

1.3 Pre-Processing

The 179 samples which were collected were pre-processed to gather insight into the data. The software that was employed was IBM SPSS Statistics 26.0. Most of the parametric statistical tests like ANOVA as well as many ML techniques are performed based on the assumption that the provided data is statistically normal. Hence test of normality was conducted for each feature. These tests took into account parameters like skewness, kurtosis and histogram plot. The results were supplemented with Kolmogorov-Smirnov (K-S) and Shapiro-Wilk (S-W) test of normality.

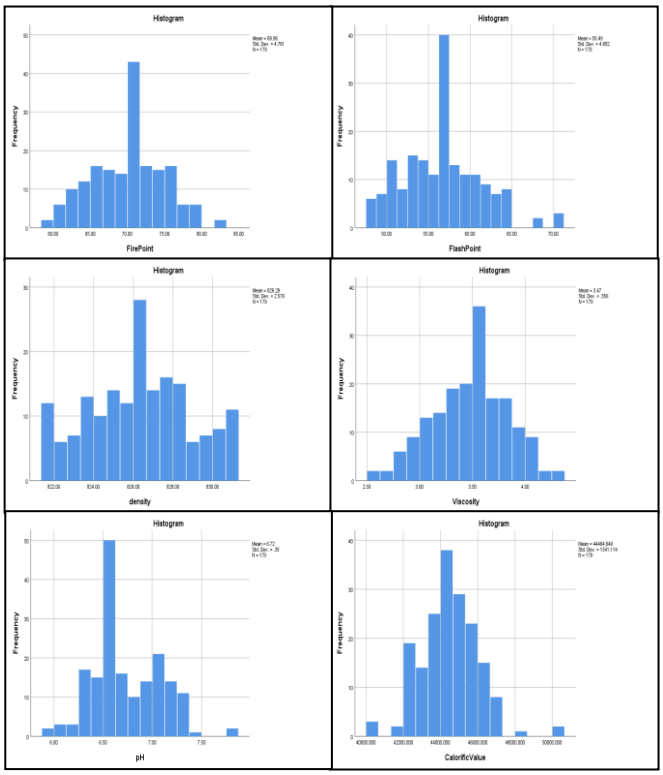

Chart 1: Histogram plot of the features

The measurements of refractive index (RI) of the samples show that it does not vary much across the samples and the range for this was found out to be 0.037 across 179 samples. As a result, the standard deviation for RI was 0.00146, which is too low and hence nearly constant. For the rest of the features, it can be seen from Chart 1. that the features are roughly normally distributed. This can be further be supported from the K-S test and S-W test from Table 3. The cut-off value for K-S and S-W significance is 0.05. If a feature has value above this, the feature follows normal distribution. Below 0.05 there will be significant deviation from normal distribution behaviour. And for the number of samples <300 the skewness and kurtosis z-value i.e skewness/std.error or kurtosis/std.error should lie between ±3.29. Though density and flash point is failing the K-S and S-W test, the ML techniques like SVM, ANN does not require a normally distributed data, hence we proceed.

Secondly, the data was analyzed for any outliers or leverage points and also tested for multicollinearity. For this, Cook's distance and correlation matrix was considered. For a particular observation to be an outlier, the Cook's distance for that observation is computed.

If, the relation is given as $Y = \beta X + \varepsilon$

Where, X is a n x p matrix of p independent variable over n observations or samples, then β is the coefficient vector explaining the influence of each Xi over dependent variable Y. ε is the error vector.

Table -2: Descriptive Statistics of features

Parameter	Mean	Variance	Std. Deviation	Skewness	Kurtosis
Flash Point	56.4947 (0.35)	21.924	4.68232	0.418 (0.182)	0.157 (0.361)
Fire Point	69.9596 (0.355)	22.668	4.76106	-0.026 (0.182)	-0.328 (0.361)
Density	826.29 (0.5985)	6.638	2.57646	-0.006 (0.182)	-0.734 (0.361)
Viscosity	3.4697 (0.0266)	0.127	0.35583	-0.17 (0.182)	-0.21 (0.361)
pH	6.725 (0.02615)	0.122	0.34987	0.48 (0.182)	-0.69 (0.361)
Calorific Value	44464.65 (115.2)	2375033	1541.114	0.258 (0.182)	1.253 (0.361)

*Bracket value denotes std. error

Table -3: Kolmogorov-Smirnov and Shapiro-Wilk test of normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Features	Statistic	df	Sig.	Statistic	df	Sig.
FlashPoint	.189	29	.010	.881	29	.004
FirePoint	.169	29	.034	.935	29	.074
density	.189	29	.009	.902	29	.011
Viscosity	110	29	.200*	.971	29	.586
pH	120	29	.200*	.959	29	.303
Calorific Value	160	29	.056	.927	29	.047

*This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Then Cook's distance is given as,

$$D_i = \frac{r_i}{p}\frac{h_{ii}}{1-h_{ii}} \qquad i = 1,2,\dots\dots,n$$

$$r_i = \frac{e_i}{\sqrt{s_e{}^2(1-h_{ii})}} \qquad i = 1,2,\dots\dots,n$$

h_{ii} is the hat matrix and s_e is the standardized sum square error. While e_i is the i^{th} error term. It is found that D_i follows F distribution and thus,

$$D_i < F_{p,n-p-1}$$

For our case, n=179, p =5,

$$D_i < F_{5,\ 173} = 1.84$$

Thus, Cooks's distance was calculated, on the basis of which 16 observations were observed to exceed this limit. Because of which these observations were identified as outliers and hence, removed from the dataset. These observations were not used to develop the correlation matrix. The feature values were min-max normalized before calculating the correlation matrix.

Table -4: Feature correlation matrix.

	pH	FlashPoint	Density	Viscosity	FirePoint
pH	1	-0.63	-0.61	0.362	0.764
FlashPoint	-0.63	1	0.412	0.18	-0.85
Density	-0.61	0.412	1	-0.528	-0.594
Viscosity	0.362	0.018	-0.528	1	0.082
FirePoint	0.764	-0.85	-0.594	0.082	1

The correlation matrix shows that the parameters are multicollinear and will require dimensional reduction. This can be done use PCA or Factor Analysis. However, with dimensional reduction valuable information is lost as it does not take into account the covariance structure between the features. Also, PCA techniques are recommended when there are large number of features. But, we have only 5 features and if we further reduce the dimensions, the models will be oversimplified causing underfitting and hence, it will not capture the desired pattern. Hence, we go ahead with all the features without dimension reduction with the assumption that ML models will handle the multicollinearity.

1.4. Prediction model

Machine learning model was developed on a system with Intel i5 processor and windows 10 operating system. R studio and Spyder are the two integrated development environments used to work on the models. R and python are the two programming languages used. SVR model is developed using R programming language and Python programming language is used to prepare code for MVR and ANN. ANN model is developed using python version 3.8.5 in Spyder environment version 4.1.5. Python code for 2-layered ANN, 3 layered ANN models are developed. While developing this model various libraries such as pandas, NumPy etc are used.

Model is developed using 2 datasets, dataset_179 (contains 179 observations) and dataset_163 (contains 163 observations). In total 3 ANN models were developed, first was built on 179 samples using 3 layered ANN. Second and third model were built on 163 samples using two layered and 3 layered ANN. There were some outliers presents in 179 datasets as found during pre-processing because of which accuracy of model was suffering. After removing those outliers' new dataset i.e., dataset containing 163 observations were created. K-fold cross-validation was carried out to improve on the generalization of model over entire dataset.

Table -5: ANN model parameters and hyperparameters.

	ANN Model 1	ANN Model 2	ANN Model 3
Dataset	dataset_179	dataset_163	dataset_163
Test set	Observations 1 to 30	Observations 1 to 30	Observations 1 to 30
Train set	Observations 30 to 179	Observations 30 to 163	Observations 30 to 163
No of layers	3	2	3
No of nodes	[6, 4, 1]	[13, 1]	[6, 3, 1]
Normalization technique	Min-max normalization	Max normalization	Max normalization
Learning rate	0.09	9	6.7
beta	0.95	-	0.3
lambda	0.0001	-	0.0001
No of iterations	14501	100000	100000

SVR model is developed by using R version 4.0.0 in R studio environment. Predefined function of SVR is available and extensive coding is not required in this case. Various libraries are used to split the dataset and import SVR predictive function. CaTools is used to split the train and test data and e1071 to use SVM function to train and test SVR model. SVR model was developed using dataset_163. Rigorous iterations were carried out by changing hyperparameters and even by changing architecture of SVR model.

MVR model is developed by using predefined functions available in Sklearn library. MVR model was developed using python version 3.8.2 in Spyder version 4.1.5. Model was built on datset_163. Developing and understanding MVR model is very easy and we do not have to perform meticulous trial and error process to find correct values of hyperparameters. Developing MVR model is very easy take less time relative to the other models.

Table -6: SVM model parameters and hyperparameters.

Parameter	value
Dataset	dataset_163
Test- train split ratio	0.8
seed	0
Kernel	radial
cost	0.98
gamma	0.17

2. Result and Discussion

In order to assess the accuracy of model R squared value and RMSE value is used. Graphs (scatter plots) are plotted between true value and predicted value for test and train dataset. Graphs allow us to visualise the deviation between true value and predicted value. Desirable shape of graph is y = x straight line.

Table -7: R^2 and RSME values of prediction model

Model name	(Test data) R^2	(Test data) RMSE (kcal/kg)
ANN Model 1	0.784	166.11
ANN Model 2	0.854	130
ANN Model 3	0.939	83.6
SVR Model	0.564	106.4
MVR Model	0.67	263.7

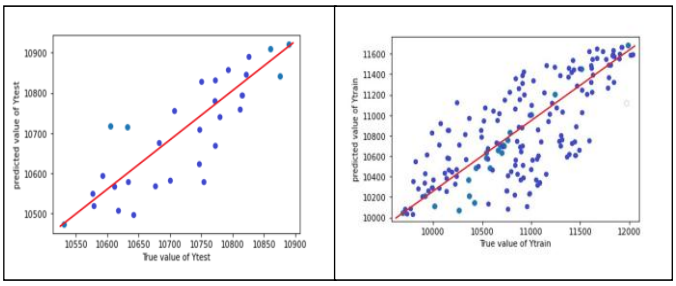

Chart 2: ANN Model 1 test train prediction plot

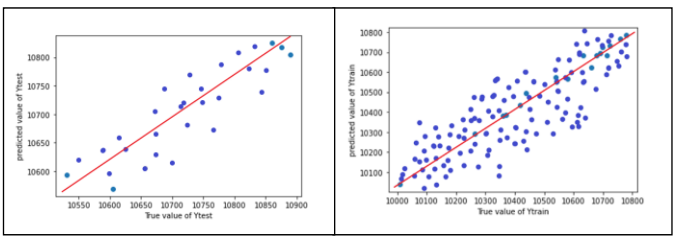

Chart 3: ANN Model 2 test train prediction plot

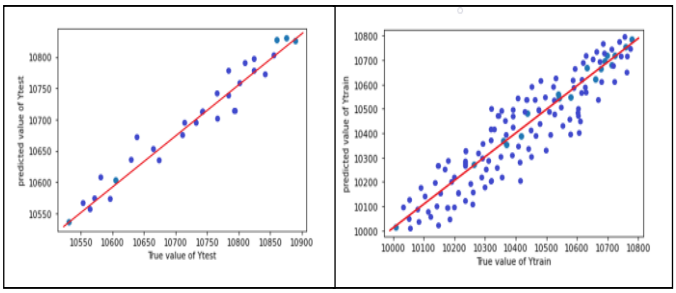

Chart 4: ANN Model 3 test train prediction plot

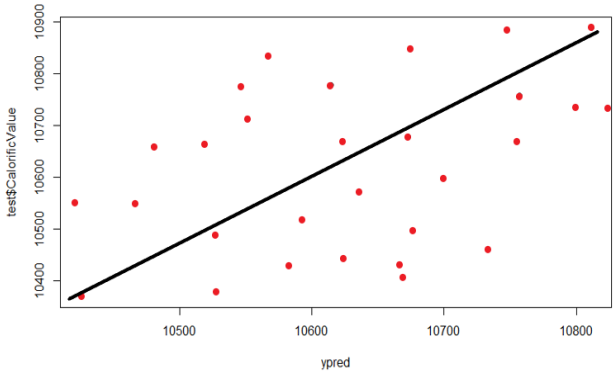

Chart 5: SVR prediction plot for test data

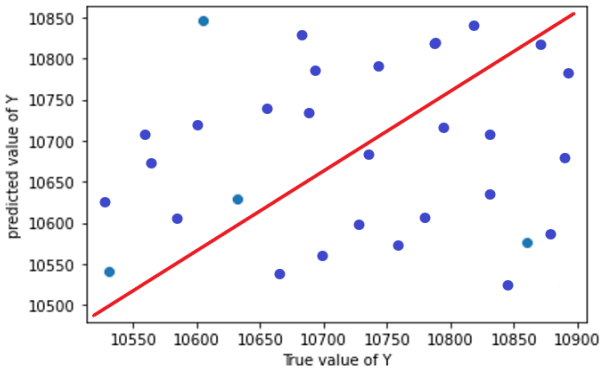

Chart 6: MVR prediction plot for test data

We did comparative study of ANN, SVR and MVR model based on how well it fits the data and prediction accuracy (for our application). Other parameters such as complexity of mathematics, easiness of hyperparameter tuning process, parameters to tune (flexibility), hardware system required, Time of execution of programme are also taken into consideration.

All these properties are quantified by giving certain weights to each of them. Points are assigned to each of the model considering their performance in each section. More weightage is given to prediction accuracy and flexibility to create more classifiers. (1 is lowest and 5 is highest)

Table 8 : Comparative study of ANN, SVR and MVR

parameters	Weights (W)	ANN		SVR		MVR	
			score		score		score
R^2	5	0.939	5	0.564	1	0.67	2
RMSE(kcal/kg)	5	83.6	5	106.4	4.5	263.7	1
No of parameters	3	8	5	3	2.5	1	1
Easiness of tuning process	3	low	1	medium	3	high	5
Complexity of maths	2	high	1	high	1	low	5
Time of execution	4	31 sec	1	0.09 sec	5	0.03 sec	5
total			74		66		63

The results in terms of R^2 and RSME values have shown that the ANN model is giving sufficient accuracy. The reason for the low accuracy of SVR and MVR models can be attributed to the fact that the features are multicollinear. Linear Kernel of SVM is very similar to Logistic Regression, and hence the effect of multicollinearity has a similar effect in the case of Linear Kernel of SVM. RBF Kernel is based on distance between the data points, similar to K-Nearest Neighbors. So It doesn't make much sense to get feature importance in this case, rather we look at data points which has influenced the decision in favor of a class to get an interpretation of the model. So, RBF kernel is also impacted by multicollinearity problem. We have to remove multicollinearity if we want to use weight vectors β directly for feature importance. While neural networks generally do not suffer from multicollinearity because they tend to be overparameterized. The extra learned weights create redundancies that make effects that influence any small subset of features (such as multicollinearity) insignificant. Multicollinearity is a problem in linear regression mainly because of process of model fitting. Assuming that there exists a unique solution to the problem, the parameters can be estimated by inverting X^TX. But this is not possible in the case of perfect collinearity as the matrix is not invertible, also it is difficult in the case of imperfect collinearity (as in our case), because the inverse is inaccurate due to its large condition number. In ANN due to its overparameterization, the coefficients or weights of a neural network are inherently difficult to interpret. However, it is this very redundancy that makes the individual weights unimportant. At each level of the network, the inputs are linear combinations of the inputs of the previous level. The final output is a function of many combinations of sigmoidal functions involving high order interactions of the original predictors. And hence a 3 layered Network is superior in handling the problems against multicollinearity compared to 2 layered model. Thus, neural networks guard against the problems of multicollinearity at the expense of interpretability.

3. CONCLUSIONS

We have used three regression models i.e., ANN, SVM, MVR to predict calorific value of diesel using basic properties such as density, viscosity, pH and others. Model 3 of ANN has highest prediction accuracy and can be used for real time prediction of calorific value of diesel. The study demonstrates that there exists a relationship between calorific value and other properties of diesel such as density, viscosity etc. which can be explained with the use of prediction models. ANN model has high prediction accuracy and is superior to other two algorithms because it is superior in handling multicollinearity. ANN is flexible but complex and tedious process whose speed can be increased by using high performance hardware system. SVR model is a fast, correlation tool but not much accurate in predicting calorific value of diesel. MVR is easy to understand and easy to implement. MVR gives a clear algebraic relationship between calorific value and other properties. But considering the case of multicollinearity, SVM and MVR are not accurate enough to correctly predict the calorific value.

REFERENCES

[1] John Nail, Diana Hamilton, "Analysis of a bomb calorimeter failure," Journal of Chemical Health and Safety, vol. 17, Issue 2, pp 21-23,2010, https://doi.org/10.1016/j.jchas.2009.08.002.

[2] W. F. Faragher, J. C. Morrell and J. L. Essex, "Relationship between Calorific Value and Other Characteristics of Residual Fuel Oils and Cracked Residuum's," Industrial and Engineering Chemistry, pp 933, 1970, https://doi.org/10.1021/ie50238a012.

[3] L. E. Oliveira and M. Da Silva, "Relationship between cetane number and calorific value of biodiesel from Tilapia visceral oil blends with mineral diesel," Renewable Energy and Power Quality, 2013.

[4] R. K. Verma, P. Suwalka and J. Yadav, "Detection of adulteration in diesel and petrol by kerosene using SPR based," Optical Fiber Technology, vol. 43, pp. 95-100, 2018, https://doi.org/10.1016/j.yofte.2018.04.011.

[5] Zhimin Lu, Xiaoxuan Chen, Shunchun Yao, Huaiqing Qin, Lifeng Zhang, Xiayang Yao, Ziyu Yu and Jidong Lu, "Feasibility study of gross calorific value, carbon content, volatile matter content and ash content of solid biomass fuel using LIBS," Fuel, vol. 258, pp.116150, 2019, https://doi.org/10.1016/j.fuel.2019.116150.

[6] C.-A. Baldrich F., L.-A. Novoa M." Rapid characterization of diesel fuel by infrared spectroscopy ", CT&F Ciencia, Tecnol. Y Futur. 3 (2006) 171–182.

[7] R.E. Morris, M.H. Hammond, J.A. Cramer, K.J. Johnson, B.C. Giordano, K.E. Kramer, et al., Rapid fuel quality surveillance through chemometric modeling of near-infra□red spectra rapid fuel quality surveillance through chemometric modeling of near infrared spectra, Energy 4 (2009) 1610–1618, http://dx.doi.org/10.1021/ef800869t

[8] M.C. Breitkreitz, I.M.J. Raimundo, J.J.R. Rohwedder, C. Pasquini, H.A. Dantas Filho, G.E. José, et al., Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration, Analyst 128 (2003) 1204–1207, http://dx.doi.org/10.1039/b305265f

[9] F.B. Gonzaga, C. Pasquini, "A low cost short wave near infrared spectrophotometer: application for determination of quality parameters of diesel fuel," Anal. Chim. Acta 670 (2010) 92–97, http://dx.doi.org/10.1016/j.aca.2010.04.060.

[10] M.A. Fahim, T.A. Alsahhaf, A. Elkilani, "Chapter2-Refinery Feedstocks and Products", Fundamentals of Petroleum Refining, Elsevier, pp. 11-31, 2010, http://dx.doi.org/10.1016/B978-0-444-52785-1.00002-4.

[11] R.M. Balabin, E.I. Lomakina, R.Z. Safieva, "Neural network (ANN) approach to biodiesel analysis: analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy," Fuel, vol 90, pp. 2007–2015, 2011, https://doi.org/10.1016/j.fuel.2010.11.038

[12] J.C.L. Alves, R.J. Poppi, "Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM)," Talanta, vol 104, pp. 155–161, 2013, http://dx.doi.org/10.1016/j.talanta.2012.11.033.

[13] V.O. Santos, F.C.C. Oliveira, D.G. Lima, A.C. Petry, E. Garcia, P.A.Z. Suarez, et al., "A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis," Analytica Chimica Acta, vol. 547, pp. 188–196, 2005, http://dx.doi.org/10.1016/j.aca.2005.05.042.

[14] J.C.L. Alves, R.J. Poppi, "Simultaneous determination of hydrocarbon renewable diesel, biodiesel and petroleum diesel contents in diesel fuel blends using near infrared (NIR) spectroscopy and chemometrics," Analyst, vol. 138, pp. 6477–6487, https://doi.org/10.1039/C3AN00883E

[15] .A. Honorato, B. de B. Neto, M.F. Pimentel, L. Stragevitch, R.K.H. Galvão, "Using principal component analysis to find the best calibration settings for simultaneous spectroscopic determination of several gasoline properties," Fuel, vol. 87 pp. 3706–3709, 2008 , http://dx.doi.org/10.1016/j.fuel.2008.06.016.

[16] R.R. Bouckaert, E. Frank, G. Holmes, D. Fletcher, "A comparison of methods for estimating prediction intervals in NIR spectroscopy: size matters," Chemometrics and Intelligent Laboratory Systems, vol 109 pp. 139–145, 2011, http://dx.doi.org/10.1016/j.chemolab.2011.08.008

[17] R.M. Balabin, R.Z. Safieva, "Near-infrared (NIR) spectroscopy for biodiesel analysis: fractional composition, iodine value, and cold filter plugging point from one vibrational spectrum," Energy Fuel, vol. 25 pp.2373–2382, 2011, http://dx.doi.org/10.1021/ef200356h.

[18] S. U. Patel, J. B. Kumar, Y. P. Badhe, B. K. Sharma, S. Saha, S. Biswas, A. Chaudhury, S. S. Tambe and B. D. Kulkarni, "Estimation of gross calorific value of coals using artificial neural networks," Fuel, vol. 86, pp.334-344, 2007, https://doi.org/10.1016/j.fuel.2006.07.036.

[19] C. D. Everard, K. P. McDonnell and C. C. Fagan, "Prediction of biomass gross calorific values using visible and near infrared spectroscopy," Biomass and Bioenergy, vol. 45, pp. 203-211, 2012, https://doi.org/10.1016/j.biombioe.2012.06.007.

[20] Q. Feng, J. Zhang, X. Zhang and S. Wen, "Proximate analysis-based prediction of gross calorific values of coals: A comparison of support vector machine, alternating conditional expectation and artificial neural network," Fuel Processing Technology, vol. 129, pp. 120-129,2014.,https://doi.org/10.1016/j.fuproc.2014.09.001

[21] L. Xu, Y. Cheng, R. Yin and Q. Zhang, "Comparative study of regression modeling methods for online coal calorifc value prediction from flame radiation features.," Fuel, vol. 142, pp. 164-172, 2014 https://doi.org/10.1016/j.fuel.2014.10.081.

[22] S. S. Matin and S. C. Chelgani, "Estimation of coal gross calorific value based on various analyses by random forest method," Fuel, vol. 177, pp. 274-278, 2016, https://doi.org/10.1016/j.fuel.2016.03.031.