PREDICTING MODELLING OF BIOOIL AND BIOCHAR PRODUCTION FROM MICROWAVE PYROLYSIS OF VARIOUS BIOMASS

Rutvik Kumar

PREDICTING MODELLING OF BIOOIL AND BIOCHAR PRODUCTION FROM MICROWAVE PYROLYSIS OF VARIOUS BIOMASSES

Thesis submitted to

Indian Institute of Technology Bhubaneswar
for the award of the degree
of

Master of Technology in Civil Engineering with specialization

in

"Environmental Engineering"

by

RUTVIK KUMAR

19CE02026

Under the guidance
of

Dr. REMYA NEELANCHERRY

SCHOOL OF INFRASTRUCTURE

INDIAN INSTITUTE OF TECHNOLOGY BHUBANESWAR
May 2024

<u>CERTIFICATE</u>

This is to certify that the thesis entitled **Predictive modelling of biooil and biochar production from microwave pyrolysis of various biomasses**, submitted by **Rutvik Kumar (19CE02026)** to Indian Institute of Technology Bhubaneswar, is a record of bonafide research work under my supervision and I consider it worthy of consideration for the award of the degree of Master of Technology in Environmental Engineering of the Institute**.**

Place: Bhubaneswar

Date: 06/05/2024

Dr. R. Neelancherry
(Supervisor)

ACKNOWLEDGMENT

DECLARATION

I certify that

a) The work contained in this thesis is original and is done by me under the guidance of my supervisors.

b) The work has not been submitted to any other Institute for any degree or diploma.

c) I have followed the guidelines provided by the Institute in preparing the thesis.

d) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e) Whenever I have materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in references.

f) Whenever I have quoted written materials from other sources, I have put them in quotation marks and given due credit to the sources by citing them and giving required details in the references.

RUTVIK KUMAR

ABSTRACT

Accurately predicting the properties of bio-char and bio-oil holds paramount importance for their effective utilization. However, these predictions are significantly swayed by the diverse compositions of biomass and the specific conditions of pyrolysis. Biomass, encompassing crops, wood, agricultural residues, and forestry remnants, stands as a pivotal renewable resource for the generation of sustainable bioenergy and eco-friendly materials. Microwave heating presents itself as a superior alternative to conventional heating due to its myriad advantages. Moreover, the judicious inclusion of suitable catalysts exerts a profound impact on the product selectivity during microwave pyrolysis. This research endeavour offers a novel perspective on forecasting the attributes of biooil and bio-char derived from various biomass compositions subjected to distinct pyrolysis conditions. It employs a range of machine learning methodologies to facilitate these predictions. Models like linear regression; random forest regression, support vector machine, adaptive boosting and gradient boosting were used to predict the biooil yield (%), SSA and HHV. Also predicted the biochar yield (%), viscosity and HHV. Various correlations and dependencies have found on the proximate analysis parameters like Ash content, volatile matter content and fixed carbon content along with ultimate analysis parameters. After doing data visualization and data processing various models have been created using all possible combinations of input parameters. For yield (%), the R^2 was up to 0.87 for biochar and 0.63 for biooil. For HHV, R^2 was 0.96 and 0.71 for biochar and biooil respectively. From experimental validation, the observed error was similar to the error of models. Proximate analysis and process parameters were turned out to be the best set for prediction with least error. Also, the tree based models were turned out to be the best models among others. By doing so, it not only enhances our understanding of bio-char and bio-oil behaviour but also contributes to the optimization of biomass-based energy production and environmentally conscious materials manufacturing.

Keywords: bio-char; machine learning; linear regression; random forest regression, support vector machine, adaptive boosting, gradient boosting

CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AC   Ash Content

FC   Fixed Carbon Content

FW   Food Waste

HHV   High Heating Value

MAE   Mean Absolute Error

MSE   Mean Squared Error

MC   Moisture Content

RT   Reaction Time

RMSE   Root Mean Squared Error

R^2 R-Squared

SWM   Solid Waste Management

SSA   Specific Surface Area

Vis   Viscosity

VM   Volatile Matter

WtE   Waste-to-energy

Chapter 1
INTRODUCTION

1.1 General

The excessive depletion of non-renewable fossil fuels has led to profound concerns regarding energy security and environmental degradation. Consequently, the focus has shifted towards harnessing biomass energy, which is abundant and sustainable. Thermo chemical conversion stands out as an efficient and cost-effective means of transforming biomass into biofuels, which can be further synthesized into desired chemicals or directly utilized. This conversion process encompasses combustion, gasification, and pyrolysis, with pyrolysis, a thermally induced decomposition method conducted in an oxygen-free environment, emerging as a promising technique due to its simplicity and rapidity of operation. Pyrolysis holds great potential for biomass conversion, offering a pathway to address energy and environmental challenges (Chowdhury et al., 2021).

Assessing the suitability of various biomass feedstocks for energy conversion processes, including combustion, pyrolysis, gasification, liquefaction, and hydrothermal processing, necessitates comprehensive characterization and experimental investigations. Among the critical properties for these thermal conversions, the HHV, also known as gross calorific value or gross energy, plays a pivotal role. HHV quantifies the heat released when a unit mass of fuel undergoes complete combustion, producing water in a condensed state (Kumar et al., 2013). Traditionally, researchers measure the HHV of biomass and other fuel sources using an adiabatic oxygen bomb calorimeter, a widely accepted, precise, but time-consuming and cost-intensive method. Unfortunately, this method isn't readily accessible to all researchers interested in HHV determination for biomass. Pyrolysis, a process involving thermal decomposition in the absence of oxygen, constitutes the initial step in generating primary products for combustion and gasification processes, followed by the total or partial oxidation of intermediate products. Biomass pyrolysis yields solid char, liquid bio-oil, and gaseous fuel, with their proportions influenced by operational variables (Zhao et at., 2014). Lower process temperatures and extended vapor residence times favour char production, while higher temperatures and longer vapor residence times enhance gas production. Optimal conditions for liquid production involve moderate temperatures and shorter vapor residence times. Much research on pyrolysis employs

conventional heating sources like electric and gas heaters. However, utilizing microwave irradiation for biomass pyrolysis shows promise due to several advantages over conventional methods (Cheng et al., 2012).

Evaluating biomass suitability for energy conversion processes necessitates characterizing properties like HHV, while pyrolysis serves as a critical step in producing biofuels. Microwave-assisted pyrolysis presents a promising avenue for its efficiency and potential benefits compared to traditional methods. In this work, a comprehensive analysis of the biooil and biochar yield, HHV and specific surface area prediction was presented. Using machine learning tools, the quality was predicted according to biomass compositions and pyrolysis conditions, and the correlation analysis was carried out. This study provides feasible thinking for the prediction of the characteristics of biooil and biochar obtained by biomass with different compositions under different pyrolysis conditions.

1.2 Objective

Objectives for this study are as mentioned below:

1. To predict the yield (%), SSA and HHV for biochar and the yield (%), Viscosity and HHV for biooil.

2. Analyses of the process parameters such as reaction temperature, reaction time and heating rate and the proximate analysis parameters such as VM (%), AC (%) and FC (%) and ultimate analysis parameters like C (%), H (%), N (%) and O (%).

3. To get minimum set of parameters from proximate or ultimate for prediction.

<div align="right">

Chapter 2
LITERATURE REVIEW

</div>

2.1 General

The use of machine learning algorithms and experimentation processes will be required for predictive modelling of bio-oil and bio-char production. In this chapter, the method for biooil and biochar production has been covered, describing the microwave pyrolysis process, the different feedstock used, and the important features considered for predicting yield, HHV, and specific surface area. Additionally, information about basic machine learning algorithms like linear regression, and some advanced algorithms like random forest, adaptive boosting regression, and artificial neural network, is provided, along with a detailed overview of how these algorithms work.

2.2 Microwave Pyrolysis

Microwave pyrolysis is an innovative and efficient technique employed in the production of bio-oil and bio-char from biomass. This method leverages microwave radiation to induce thermal decomposition of biomass in the absence of oxygen, leading to the creation of valuable products.

2.2.1 Principle of Microwave Pyrolysis

Microwave pyrolysis relies on the principle of selective heating. When microwave radiation is applied to biomass, water molecules within the material absorb the microwave energy and rapidly heat up, leading to the formation of steam. As the temperature increases, the biomass undergoes pyrolysis, breaking down into its constituent components: solid bio-char, liquid bio-oil, and gaseous products.

2.2.2   Advantages of Microwave Pyrolysis

1. Speed: Microwave pyrolysis is significantly faster compared to conventional heating methods. The selective heating of water molecules within the biomass allows for rapid and uniform heating throughout the material.

2. Energy Efficiency: It is more energy-efficient because it directly targets the water content within the biomass, reducing energy wastage.

3. Enhanced Product Quality: The controlled and uniform heating in microwave pyrolysis can lead to higher yields of bio-oil and bio-char with improved quality.

4. Reduced Environmental Impact: Microwave pyrolysis can be carried out in an oxygen-free environment, reducing the production of harmful by-products and minimizing emissions.

2.2.3 Process and Products

In a typical microwave pyrolysis process:

1. Biomass is loaded into a microwave-reactive vessel.
2. Microwaves are applied, causing rapid heating of the biomass.
3. As the temperature increases, the biomass undergoes pyrolysis, releasing volatile compounds that can be condensed into bio-oil.
4. The remaining solid residue is bio-char, which is rich in carbon and can be used as a valuable soil conditioner or for carbon sequestration.
5. Gaseous by-products can also be collected and used for energy generation or other applications.

2.2.4 Applications

- Bio-Oil: The liquid bio-oil obtained from microwave pyrolysis can be a valuable feedstock for the production of biofuels, chemicals, and other high-value products. Its composition can vary depending on the biomass source and pyrolysis conditions.

- Bio-Char: Bio-char, the solid residue, finds applications in agriculture as a soil amendment to improve soil fertility, water retention, and carbon sequestration. It can also be used as a precursor for activated carbon production.

Microwave pyrolysis is a promising method for the sustainable production of bio-oil and biochar from biomass. Its advantages in terms of speed, energy efficiency, and product quality make it an attractive option for biomass conversion, contributing to the development of renewable energy and environmentally friendly materials.

2.3 Biomass feedstocks dataset

The experimental data of bio-oil and biochar were collected from the literature. The dataset consists of a wide range of biomass feedstocks such as commercial fuels, industrial wastes, forest wastes (including branches and leaves), energy crops and cereals, which all can be used to build

a model valid in a wide range of biomass feedstocks to predict the yield, SSA and HHV. Table 2.1 gives the sample dataset of process parameters from the original dataset which were used for prediction.

Table 2.1.Sample Dataset of Process Parameters and target parameters

biomass type	C (%)	H (%)	N (%)	O (%)	VM (%)	Ash (%)	FC (%)	T (°C)	RT (min)	HR (°C/ min)	Yield (%)	SSA (m²/g)	HHV (MJ/KG)
Durian Wood	33.77	11.40	6.06	48.77	85.80	10.52	3.68	350	180	1.94		2.57	14.31
Durian Wood	33.77	11.40	6.06	48.77	85.80	10.52	3.68	450	180	2.5		45.68	14.31
Durian Wood	33.77	11.40	6.06	48.77	85.80	10.52	3.68	550	180	3.05		221.00	14.31
marine macroalgae	39.78	6.35	5.40	48.46		24.51		200	120	1.66	78.34	1.91	0.00
marine macroalgae	39.78	6.35	5.40	48.46		24.51		400	120	3.33	63.64	70.29	0.00
marine macroalgae	39.78	6.35	5.40	48.46		24.51		600	120	5	37.96	61.81	0.00
empty fruit bunch (EFB)	53.04	6.39	0.87	39.37	82.07	4.88	13.05	500	60	8.33		1.41	16.66
empty fruit bunch (EFB)	53.04	6.39	0.87	39.37	82.07	4.88	13.05	300	120	2.5		1.42	16.66
empty fruit bunch (EFB)	53.04	6.39	0.87	39.37	82.07	4.88	13.05	700	120	5.83		1.13	16.66
rice husk (RH)	52.85	6.19	0.44	40.22	70.78	12.00	17.22	500	60	8.33		7.84	16.23
rice husk (RH)	52.85	6.19	0.44	40.22	70.78	12.00	17.22	300	120	2.5		0.39	16.23
corn straw pellets	40.02	6.01	0.88	52.86	72.29	10.02	17.68	400	30	13.3	50.17	4.26	16.61
corn straw pellets	40.02	6.01	0.88	52.86	72.29	10.02	17.68	450	30	15	49.57	12.61	16.61
Saccharina japonica algae	40.18	5.55	2.00	51.39				300	60	5		1.36	0.00
Saccharina japonica algae	40.18	5.55	2.00	51.39				450	60	7.5		210.00	0.00
Saccharina na	40.18	5.55	2.00	51.39				600	60	10		266.00	0.00

biomass type	VM	Ash	FC	C%	H%	O%	N%	PS	HR(°C/min)	RT(min)	T(°C)	Yield(%)	Vis	HHV(MJ/KG)
japonica algae														
rice husk (RH)	40.65	5.95	1.07	52.19		16.53			400	60	6.66	60.57	5.49	0.00
rice husk (RH)	40.65	5.95	1.07	52.19		16.53			700	60	11.6	52.58	7.12	0.00
sewage sludge (SS)	45.74	10.76	7.04	34.91	49.36	42.02	8.63		400	60	6.66	53.84	4.39	10.26
sewage sludge (SS)	45.74	10.76	7.04	34.91	49.36	42.02	8.63		700	60	11.6	40.35	10.71	10.26
bamboo sawdust (BS)	49.51	3.42	0.77	46.02		2.04			400	60	6.66	44.51	2.78	0.00
Raw maize straw	42.19	7.21	1.28	49.19		5.31			250	60	4.16		5.20	0.00
Raw maize straw	42.19	7.21	1.28	49.19		5.31			300	60	5	36.70	7.80	0.00

Table 2.2. Sample Dataset of Process Parameters and Target Parameters

biomass type	VM	Ash	FC	C%	H%	O%	N%	PS	HR(°C/min)	RT(min)	T(°C)	Yield(%)	Vis	HHV(MJ/KG)
Mahua seed	77.05	2.08	12.63	52.48	8.69	34.86	2.62	1	100	5	500	51.2	0.08256	29.71
Mahua seed	77.05	2.08	12.63	52.48	8.69	34.86	2.62	1	100	5	500	50.15	0.02533	35.92
Date seed	81.15	5.17	7.63	70.91	10.5	16.59	1.69	0.1	16	31.25	500	68.5	0.00142	29.06
Syrup waste	67.24	10.13	14.07	67.23	7.91	22.13	2.5	0.1	16	31.25	500	58.6	0.00131	19.29
Bamboo torrefaction	75.19	2.52	17.69	49.23	2.95	47.5	0.32	3.35	22	11.36	250	12	0.00179	
Bamboo torrefaction	75.19	2.52	17.69	49.23	2.95	47.5	0.32	3.35	22	13.64	300	28	0.00207	
Bamboo torrefaction	75.19	2.52	17.69	49.23	2.95	47.5	0.32	3.35	22	15.91	350	40	0.00364	
Napier Grass Stem (NGS)	81.5	1.75	16.7	48.6	6.01	44.1	0.99	2	30	15	450	27.7	0.00245	25.30

Napier Grass Stem (NGS)	81.5	1.75	16.7	48.6	6.01	44.1	0.99	2	30	18.33	550	30.5	0.00267	25.40
Switchgrass	84.4	3.9	11.9	42	6.1	47.4	0.4	2	6	100	600	37	0.0092	36.30
Rice husks	62.7	12.7	15.1	37.86	5.24	35.32	0.68	1.5	180	3.06	550	48.1	0.00857	18.07
Cassia siamea seed	84	5.8	0.5	40.8	6.1	48.9	3.9	1.5	50	11	550	39.86	0.337	30.18
Bael shell	91.42	5.64	2.94	40.67	6.23	41.6	1.34	1	150	3.33	500	36.23	0.00368	19.40
Date stones	79.88	1.47	18.65	48.43	6.44	42.2	0.67	0.25	10	50	500	25.67	0.01292	37.38
Hornbeam sawdust	78.14	0.45	12.63	45.18	6.59	38.69	0.05	5	30	18.33	550	29.83	0.0008	
Tunisian AS		2.71		45.64	6.19	45.43	0.5	0.84	10	43.7	437	63	0.8368	27.40
Spanish AS		0.55		50.5	6.58	42.65	0.21	0.84	10	43.7	437	40	0.0019	14.10

2.3 Machine Learning Algorithms

2.3.1 Linear Regression

Linear regression may be defined as the statistical model that analyzes the linear relationship between a dependent variable with given set of independent variables. Linear relationship between variables means that when the value of one or more independent variables will change (increase or decrease), the value of dependent variable also change accordingly (increase or decrease).

Mathematically, the relationship can be represented per equation 1:

$$Y = mX + b \tag{1}$$

Where, Y is the dependent variable (to be predicted)

X is the independent variable used to make predictions.

m is the slope of the regression line which represents the effect X on Y

b is a constant, known as the Y-intercept. If $X = 0$, Y would be equal to b.

In case the best fit line passes through origin, the equation becomes

$$Y = mX \tag{2}$$

Generalized linear regression model for any response (y) and its predictors(x_1, x_2, ...,x_n), i.e., biodiesel production process parameters in this case can be stated as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_n x_n \tag{3}$$

2.3.2 Decision Tree

A Decision Tree Regressor is a machine learning algorithm used for regression tasks. It works by recursively partitioning the data into subsets based on the values of input features to predict continuous output values. Each split in the tree represents a decision or a rule, and the outcome is obtained by averaging the target values of data points within a leaf node. The tree-building process begins with the entire dataset and aims to find the feature and split point that best separates the data based on a chosen criterion often mean squared error. This process continues recursively until a stopping criterion is met, such as a maximum depth of the tree or a minimum number of data points in a leaf node.

Decision trees are intuitive and easy to interpret, allowing users to understand the model's decision-making process. However, they can be prone to overfitting, meaning they capture noise in the training data. Techniques like pruning and setting appropriate hyperparameters help mitigate overfitting. Decision Tree Regressors are useful for modelling both linear and nonlinear relationships between input features and target variables. They are the building blocks for more complex ensemble methods like Random Forests and Gradient Boosting, which combine multiple decision trees to improve predictive accuracy. Decision trees are valuable tools in various fields, including finance, healthcare, and environmental modelling, where interpretable and accurate regression models are required.

2.3.3 Support Vector Regressor

A Support Vector Regressor is a machine learning algorithm used for regression tasks. It is based on the same principles as the Support Vector Machine, which is primarily used for classification. SVR, however, is designed to predict continuous values, making it suitable for tasks such as stock price forecasting, house price prediction, and more. SVR works by finding a hyperplane that best fits the data, while also minimizing the margin of error. In SVR, this margin is called the "tube," and the goal is to fit as many data points as possible within this tube while limiting violations (data points that fall outside the tube). The data

points that touch the edges of the tube are the support vectors, and they play a critical role in determining the hyperplane.

To achieve this, SVR uses a mathematical optimization technique that involves minimizing the L2 norm of the weight vector (hyperplane) while subject to the constraint that the error on each data point is within a specified epsilon (ε). The choice of kernel function, such as linear, polynomial, or radial basis function, allows SVR to handle nonlinear relationships between input features and the target variable.SVR is a powerful regression algorithm that uses support vectors to find the best-fitting hyperplane and can handle both linear and nonlinear relationships in the data, making it a valuable tool in predictive modelling.

2.3.4 Random Forest Regression

The Random Forest is based on applying bagging to decision trees, with one important extension: in addition to sampling the records, the algorithm also samples the variables. To determine how to create a sub partition of a partition A, the algorithm chooses the variable and split point by minimising a criterion such as Gini impurity with random forests, the choice of variable is limited to a random subset of variables at each stage of the algorithm. Compared to the basic tree algorithm, the random forest algorithm adds two more steps: the bagging and the bootstrap sampling of variables at each split:

1. Take a bootstrap (with replacement) subsample from the *records*.
2. For the first split, sample $p < P$ *variables* at random without replacement.
3. For each of the sampled variables $X_{j1}, X_{j2}, ..., X_{jp}$, apply the splitting algorithm:
   a. For each value s_{jk} of X_{jk}:
      i. Split the records in partition A, with $X_{j(k)} < s_{j(k)}$ as one partition and the remaining records where $X_{jk} \geq s_{jk}$ as another partition.
      ii. Measure the homogeneity of classes within each sub partition of A.
   b. Select the values of s_{jk} that produces maximum within-partition homogeneity of class.
4. Select the variable X_{jk} and the split value s_{jk} that produces maximum with in partition homogeneity of class.
5. Proceed to the next split and repeat the previous steps, step 2 onwards.
6. Continue with additional splits, following the same procedure until the tree is grown.
7. Go back to step 1, take another bootstrap subsample, and start the process over again.

2.3.5 Adaptive Boosting Regressor

Adaptive boosting, also known as AdaBoost, is a sequential ensemble technique that develops several weak learners using different training subsets drawn at random from the original training dataset. Weights are assigned during each training and are used when learning each hypothesis. The weights are used to compute the hypothesis error on the dataset and act as an indicator of the relative importance of each instance. After each iteration, the weights are recalculated so that instances incorrectly classified by the previous hypothesis receive higher weights. This allows the algorithm to concentrate on more difficult-to-learn instances. The algorithm's most important task is to assign revised weights to incorrectly classified instances. In regression, unlike in classification, the instances are not correct or incorrect; rather, they constitute area-value error. The computed error can be labelled as an error or not error by comparing it to a predefined threshold prediction error, and thus the AdaBoost classifier can be used. Instances with larger errors on previous learners are more likely (i.e., have a higher probability) of being chosen to train the next base learner. Finally, the weighted average or median of the individual base learner predictions is used to provide an ensemble prediction.

2.3.6 Gradient Boosting Regressor

A Gradient Boosting Regressor is a powerful machine learning technique used for regression tasks. It is an ensemble learning method that combines the predictions of multiple weaker models, typically decision trees, to create a robust and accurate predictive model.

Here's how Gradient Boosting Regressor works:

1. Sequential Learning: Gradient boosting builds an ensemble of decision trees sequentially. Each tree corrects the errors made by the previous ones. The process starts with an initial, simple model, and subsequent trees are trained to minimize the residual errors of the ensemble.

2. Gradient Descent: The term "gradient" in gradient boosting refers to the use of gradient descent optimization to minimize the loss function. It adjusts the weights (or parameters) of the individual trees in each iteration to minimize the prediction errors.

3. Weak Learners: Gradient boosting typically employs shallow decision trees as weak learners. These trees are often referred to as "stumps" and are combined to

create a strong predictive model. The depth of these trees is a hyperparameter that can be tuned.

4. Regularization: To prevent overfitting, gradient boosting incorporates regularization techniques, such as tree depth control, learning rate adjustment, and feature subsampling

5. Predictive Power: The final prediction is a weighted sum of predictions from all the weak learners, where the weights are determined during training. This ensemble model can capture both linear and nonlinear relationships in the data, resulting in a highly accurate regression model.

Gradient Boosting Regressor is widely used in various fields, including finance, healthcare, and natural language processing. It is known for its ability to handle complex datasets, robustness against outliers, and excellent predictive performance, making it a popular choice for regression tasks when high accuracy is required.

2.3.7 Predictive Model Performance Evaluation Metrics

The residue, ε_i between the ith original value, y_i and predicted value, y_0 is calculated as,

$$\varepsilon_i = y_i - y_0 \tag{4}$$

The coefficient of determination, R^2 is computed using the y_i, y_0 and the mean of the dataset, y_m.

$$R^2 = 1 - \frac{\sum(y^\circ - y^\circ i)^2}{\sum(y^\circ - y^\circ m)^2} \tag{5}$$

The mean-absolute-error is computed based on the number of the samples, n as

$$MAE = \sum \frac{|yi - y0i|}{n} \tag{6}$$

The mean-squared-error, MSE, is computed as

$$MAE = \sum \frac{(yi - y0i)^2}{n} \tag{7}$$

The root-mean-squared-error, RMSE, is computed as

$$RMSE = (\frac{\sum(yi - y0i)^2}{n})^{0.5} \tag{8}$$

The maximum error is computed as

$$\textbf{Max. Error } (y, y_0) = \max. (|y_i - y_{0i}|) \tag{9}$$

The median error is computed as

$$\textbf{Med. Error } (y, y_0) = \text{median } (|y_i - y_{01}|, \dots, |y_n - y_{0n}|) \tag{10}$$

Chapter 3
METHODOLOGY

3.1 General

In this research endeavour, the primary focus revolves around the meticulous analysis of experimental data to discern underlying patterns and trends. The data in question holds the key to understanding and potentially predicting the bio-oil and bio-char production processes. To harness the full potential of this data, a crucial preliminary step involves data pre-treatment to render it suitable for utilization within predictive models. The foundation of this research rests on the utilization of cutting-edge Machine Learning algorithms, which serve as the cornerstone of our predictive endeavours. These algorithms are strategically applied to predict the outcomes of the bio-oil and bio-char production processes. The predictive models are designed to consider various process parameters, which play a pivotal role in determining the final outcomes. Among the ensemble of Machine Learning algorithms at our disposal, we will employ linear regression and random forest regression. These techniques have been chosen based on careful observation and analysis of the experimental data. Their application is rooted in the belief that they possess the capability to effectively capture and extrapolate the intricate relationships between the process parameters and the production responses. To gauge the effectiveness and accuracy of these predictive models, several error metrics are employed. These metrics serve as benchmarks against which the performance of the Machine Learning regression models is evaluated. By quantifying the disparities between predicted values and actual outcomes, these error metrics provide invaluable insights into the predictive power and reliability of the models. The significance of this research extends beyond its immediate objectives. It holds the potential to revolutionize the bio-oil and bio-char production processes. The ability to predict these processes with a high degree of accuracy can pave the way for enhanced efficiency, reduced costs, and environmentally friendly production methods. Moreover, the insights gleaned from this research can be extrapolated to other domains, demonstrating the broader applicability of Machine Learning in industrial and scientific endeavours. In conclusion, this research embarks on a journey into the realm of data-driven analysis and prediction in the context of bio-oil and biochar production processes. Ultimately, this research has the potential to reshape the landscape of biofuel production and contribute to a more sustainable and eco-conscious future.

3.2 Prediction of Biochar yield (%), SSA and HHV

The objective is to forecast the yield (%), SSA, and HHV of Biochar using a minimal set of parameters. This prediction task is divided into three segments, one for each target variable (yield, SSA, HHV). Furthermore, within each segment, there are two approaches: (a) using proximate analysis and process parameters and (b) utilizing ultimate analysis and process parameters. This results in a total of 12 distinct models, each tailored to predict one of the three properties, depending on the chosen parameter set.

3.3 Prediction of Biooil yield (%), Vis and HHV

The goal is to predict the yield (%), Vis, and HHV of Biooil using a concise set of parameters. This prediction objective is segmented into three parts, corresponding to each target variable (yield, Vis, HHV). Additionally, within each segment, two methods are employed: (a) employing proximate analysis and process parameters, and (b) utilizing ultimate analysis and process parameters. This leads to a total of 12 unique models, each designed to forecast one of the three properties based on the selected parameter set.

Chapter 4

RESULTS AND DISCUSSION

4.1 General

Evaluating the performance of different experimentation with model is must to get the best one for prediction. This chapter includes the information about how much these parameters are correlated with each other and the performance of different machine learning algorithms using this process parameters for predicting the bio-char yield, SSA and HHV and Also for biooil yield, Vis and HHV. This chapter also includes the best process parameters to predict the accurate target variable.

4.2 Modelling for Biochar

The prediction challenge is segmented into three parts, each corresponding to a target variable (yield, SSA, HHV). Moreover, within each segment, two methodologies are applied: (a) employing proximate analysis and process parameters, and (b) utilizing ultimate analysis and process parameters. This yields a total of 12 unique models, each customized to forecast one of the three properties, based on the selected parameter set.

4.2.1 Effect of Process Parameters

Figure 4.1 and Table 2.1 serve as visual representations of our training data, showcasing scatter plots that reveal the relationships between various features and their impact on the yield (%), SSA, and HHV. The scatter plots, thoughtfully organized, exhibit how each feature's variation corresponds to these critical variables. A significant observation can be made when examining the histograms displayed along the diagonal of the scatter plots. These histograms indicate that the distribution of our features closely resembles a normal distribution. This normality in feature distribution is a crucial assumption in many statistical analyses and machine learning algorithms, enhancing the reliability of our results.

When focusing on the relationships between the yield (%) and our features, it becomes apparent that C (%), H (%), N (%), and O (%) demonstrate a positive correlation with yield (%). In contrast, SSA exhibits a positive association with Ash (%), FC (%), and T (°C). Moreover, HHV displays a positive relationship with RT (min) and HR (°C/min). The yield histogram emphasizes that the central tendency of all experimentally determined yields

14

hovers around 40%. This statistic provides a pivotal insight into the overall performance and characteristics of our dataset, serving as a reference point for further analysis and modelling.
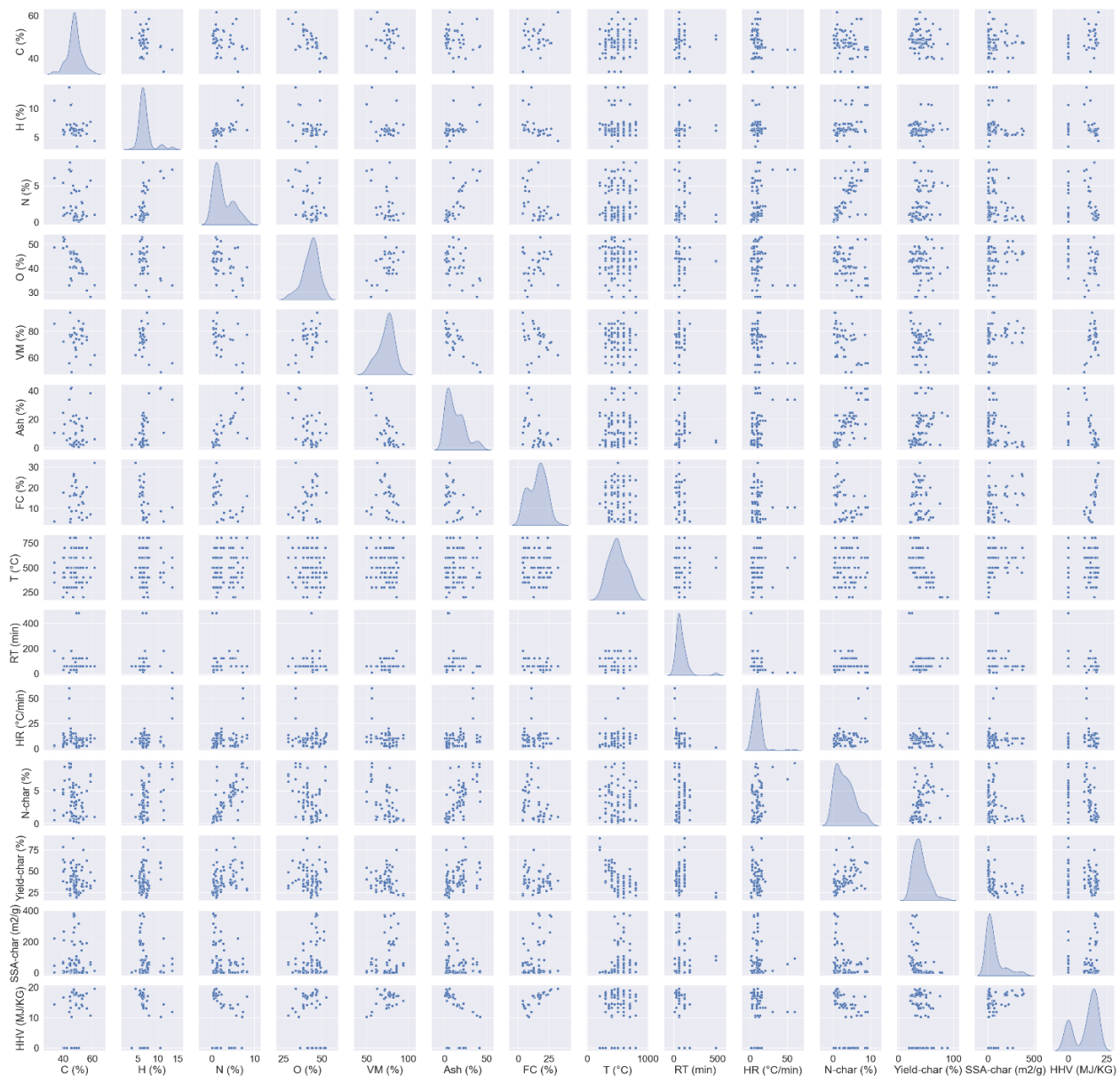
Figure 4.1 Scatter plot of the various process parameters and Target Variable.

Figure 4.2 delves into the intricate relationship between various process parameters and the target variable. A notable observation here is that the process parameters exhibit negligible correlation, confirming the absence of multicollinearity. Multicollinearity, if present, can undermine the statistical power of regression analyses, and its absence in this context is a favourable characteristic. Yield (%) exhibits strong positive correlations with H (%), N (%), and Ash (%) with correlation coefficients of 0.16, 0.37, and 0.4, respectively. Conversely, it

15

demonstrates a substantial negative correlation with T (℃), registering a correlation coefficient of -0.61. SSA shows robust positive correlations with VM (%), FC (%), and T (℃), boasting correlation coefficients of 0.14, 0.19, and 0.36, respectively. Conversely, it exhibits noteworthy negative correlations with N (%) and Ash (%) with correlation coefficients of -0.21 and -0.25, respectively. HHV, in its relationship with the process parameters, displays significant positive correlations with C (%), VM (%), and FC (%), yielding correlation coefficients of 0.31, 0.45, and 0.79, respectively. Conversely, it showcases substantial negative correlations with N (%), Ash (%), and RT (min), featuring correlation coefficients of -0.39, -0.43, and -0.33, respectively.

Despite these correlations, the scatter plots in Figure 4.2 lack a discernible trend, suggesting that an exhaustive experimental analysis of various process parameter combinations may be necessary. This underscores the pivotal role of advanced machine learning algorithms in constructing predictive models capable of quantifying yield as a function of these multifaceted process parameters.

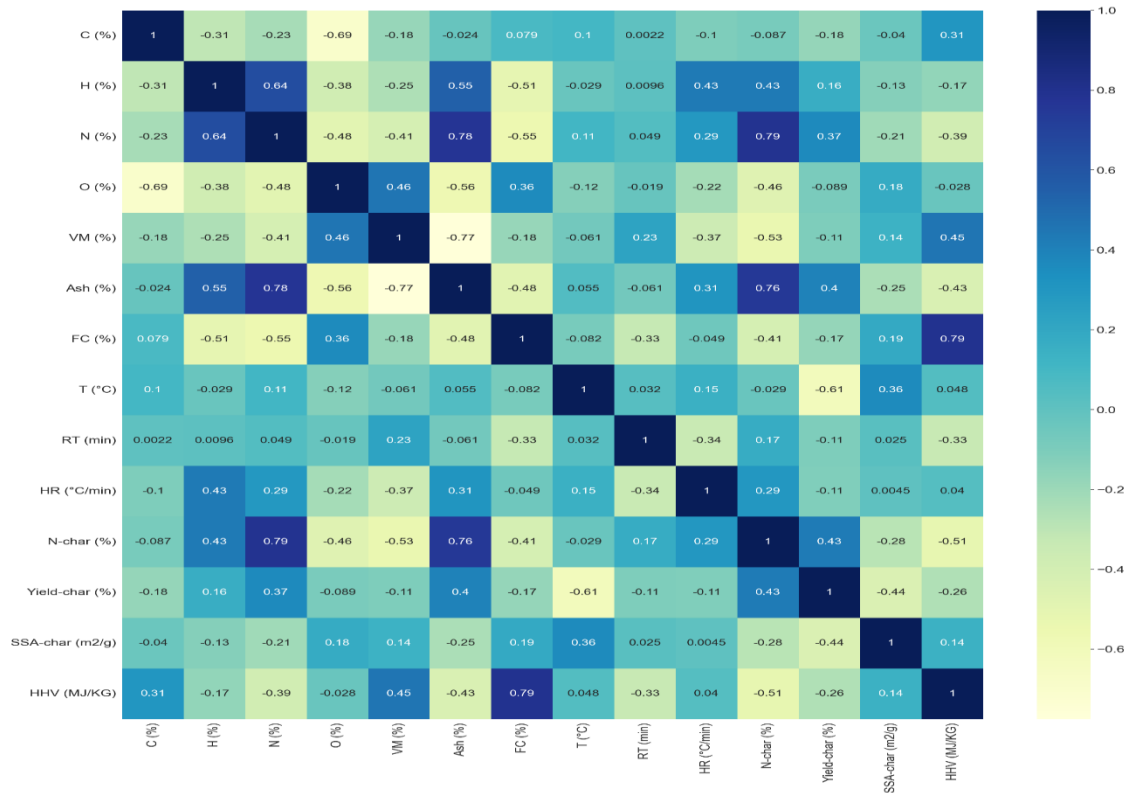|	C (%)	H (%)	N (%)	O (%)	VM (%)	Ash (%)	FC (%)	T (°C)	RT (min)	HR (°C/min)	N-char (%)	Yield-char (%)	SSA-char (m2/g)	HHV (MJ/KG)
C (%)	1	-0.31	-0.23	-0.69	-0.18	-0.024	0.079	0.1	0.0022	-0.1	-0.087	-0.18	-0.04	0.31
H (%)	-0.31	1	0.64	-0.38	-0.25	0.55	-0.51	-0.029	0.0096	0.43	0.43	0.16	-0.13	-0.17
N (%)	-0.23	0.64	1	-0.48	-0.41	0.78	-0.55	0.11	0.049	0.29	0.79	0.37	-0.21	-0.39
O (%)	-0.69	-0.38	-0.48	1	0.46	-0.56	0.36	-0.12	-0.019	-0.22	-0.46	-0.089	0.18	-0.028
VM (%)	-0.18	-0.25	-0.41	0.46	1	-0.77	-0.18	-0.061	0.23	-0.37	-0.53	-0.11	0.14	0.45
Ash (%)	-0.024	0.55	0.78	-0.56	-0.77	1	-0.48	0.055	-0.061	0.31	0.76	0.4	-0.25	-0.43
FC (%)	0.079	-0.51	-0.55	0.36	-0.18	-0.48	1	-0.082	-0.33	-0.049	-0.41	-0.17	0.19	0.79
T (°C)	0.1	-0.029	0.11	-0.12	-0.061	0.055	-0.082	1	0.032	0.15	-0.029	-0.61	0.36	0.048
RT (min)	0.0022	0.0096	0.049	-0.019	0.23	-0.061	-0.33	0.032	1	-0.34	0.17	-0.11	0.025	-0.33
HR (°C/min)	-0.1	0.43	0.29	-0.22	-0.37	0.31	-0.049	0.15	-0.34	1	0.29	-0.11	0.0045	0.04
N-char (%)	-0.087	0.43	0.79	-0.46	-0.53	0.76	-0.41	-0.029	0.17	0.29	1	0.43	-0.28	-0.51
Yield-char (%)	-0.18	0.16	0.37	-0.089	-0.11	0.4	-0.17	-0.61	-0.11	-0.11	0.43	1	-0.44	-0.26
SSA-char (m2/g)	-0.04	-0.13	-0.21	0.18	0.14	-0.25	0.19	0.36	0.025	0.0045	-0.28	-0.44	1	0.14
HHV (MJ/KG)	0.31	-0.17	-0.39	-0.028	0.45	-0.43	0.79	0.048	-0.33	0.04	-0.51	-0.26	0.14	1

Figure 4.2 Correlation between the various process parameters and Target Variable

4.2.2 Comparison of Various ML Predictive Models

The evaluation and comparison of machine learning predictive models will be executed independently for each of the three distinct segments. This assessment will be conducted through the utilization of regression plots (regplots), graphically illustrating the relationships between actual values and predicted values. Moreover, a comprehensive model evaluation will be performed, incorporating essential metrics including MAE, MSE, RMSE, and the coefficient of determination R^2. The model selection process will be based on the minimization of prediction errors (MAE, MSE, and RMSE) and the maximization of the goodness-of-fit (R^2). The chosen model, exhibiting the least error and the highest R^2 value, will be subject to hyperparameter tuning to further enhance predictive accuracy. Hyperparameter optimization is a crucial step in refining the model's performance.

Subsequently, after achieving the final optimized model, an analysis of feature importance will be conducted. This analysis will be represented in the form of a bar graph, showcasing the relative significance of each feature in the prediction process. Identifying feature importance is instrumental in understanding the variables that have the most substantial influence on the predictive model's outcomes, providing valuable insights for model interpretability and feature selection.

4.2.2.1 Prediction of Yield (%)

Based on the training dataset, predictive models have been constructed to estimate the yield percentage (%), considering ultimate analysis, proximate analysis, and process parameters. In Figure 4.3, the performance of these predictive models is evaluated using a test dataset that relies on proximate analysis and process parameters. Remarkably, the Support Vector Regressor exhibited the least favourable predictive performance, as indicated by its predicted line. In contrast, the Gradient Boosting and Random Forest models emerged as the most effective models within this dataset evaluation. Their predictive lines closely resemble the ideal scenario where the predicted values match the actual values (Y=X), highlighting their superior performance and predictive accuracy.

This outcome underscores the significance of model selection, as the choice of an appropriate algorithm can significantly impact the quality of predictions. In this context, Gradient Boosting and Random Forest models have demonstrated their potential to provide precise and reliable yield predictions in relation to the specified parameters, offering promising avenues for further analysis and application.
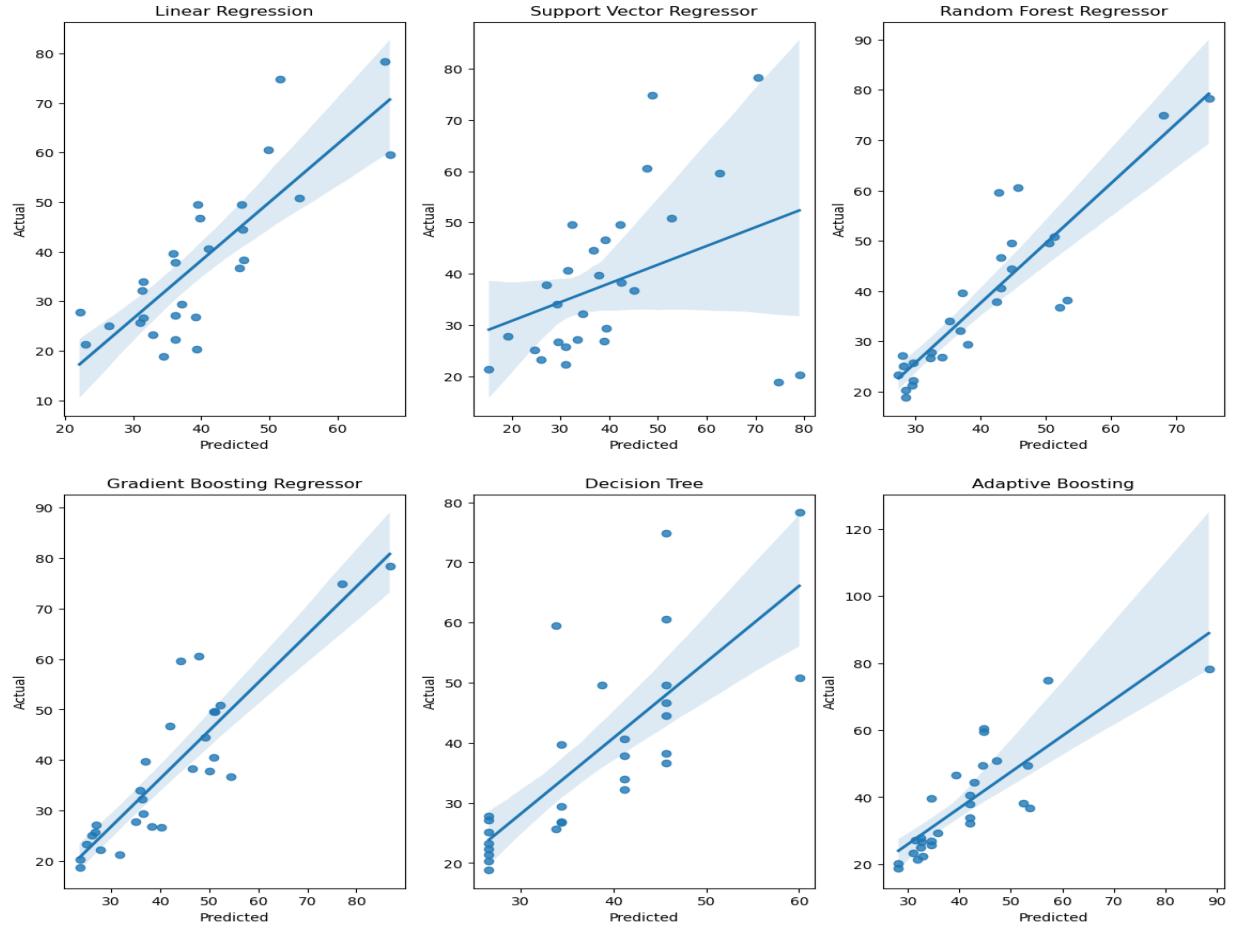
Figure 4.3 Regression plot for Yield (%) using proximate and process parameters

Table 4.1 provides a comprehensive overview of model performance as assessed through various evaluation metrics. Notably, the Support Vector Regressor stands out with the highest error metrics, boasting a MAE of 11.09 and a notably R^2 of 0.48 in comparison to all other models. Conversely, the Gradient Boosting and Random Forest models shine as the top-performing options within this dataset, characterized by their exceptional predictive accuracy. The MAE for Gradient Boosting is reported at 6.33, along with an R^2 of 0.87, demonstrating its robust performance. The Random Forest model achieves an even lower MAE of 6.01 and an impressive R^2 of 0.81.

This analysis conclusively establishes that when employing proximate analysis and process parameters, the model offering the least prediction error, with an MAE of 6.31 and the highest R^2 value of 0.87, is the gradient boosting model. These results underscore the efficacy of this model in providing accurate yield predictions based on the specified input parameters.

Table 4.1.Error value for yield (%) prediction using proximate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	7.56	11.09	6.01	6.33	7.65	7.5
MSE	88.58	313.49	57.57	63.73	105.8	78.69
RMSE	9.41	17.71	7.59	7.98	10.29	8.87
R^2	0.49	0.48	0.81	0.87	0.39	0.59

In Figure 4.4, the performance of machine learning models on a test dataset, utilizing ultimate analysis and process parameters, is presented. Notably, all models displayed commendable performance on this dataset. However, it's worth noting that the Support Vector Regressor, while not performing as well as the other models, exhibited superior predictive capabilities compared to its performance with proximate and process parameters. The standout performers in this evaluation were the Gradient Boosting and Random Forest models. These models excelled, with a significant number of data points closely aligning with the ideal Y=X line, reinforcing their effectiveness and accuracy in predicting the target variable within this specific dataset.
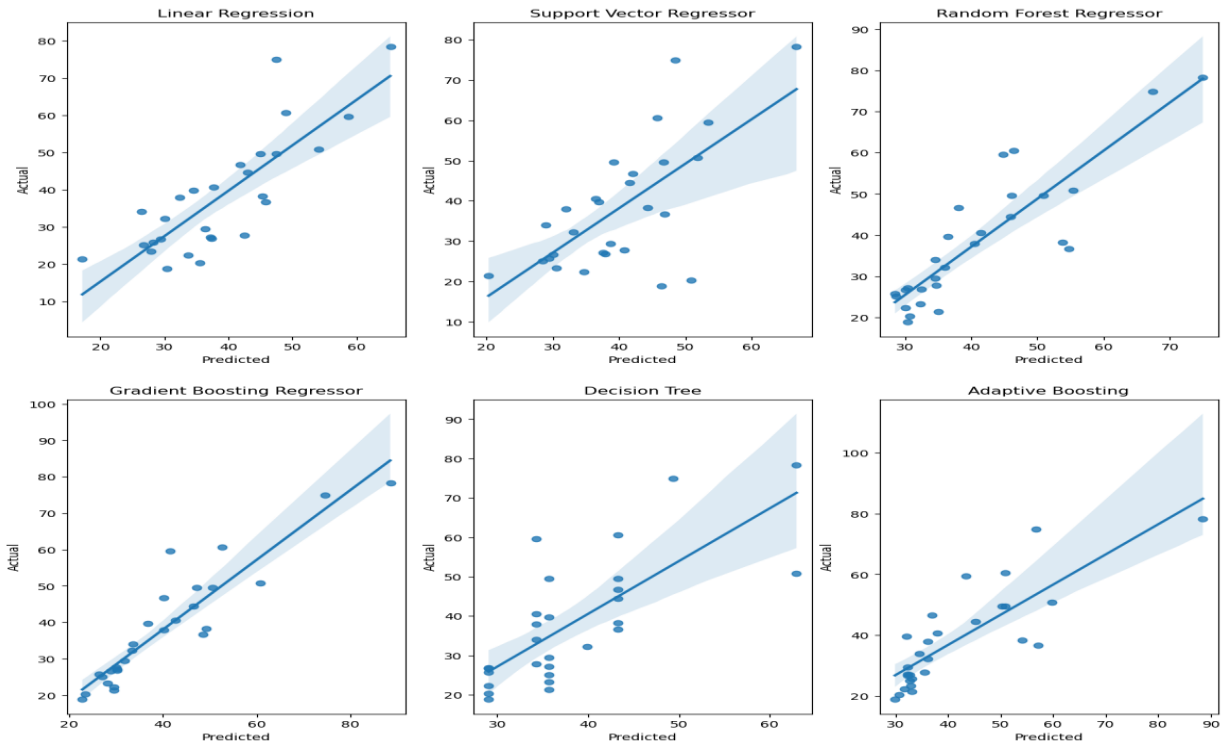

Figure 4.4 Regression plot for Yield (%) using ultimate and process parameters

Table 4.2 presents an extensive analysis of model performance utilizing evaluation metrics. The Support Vector Regressor stands out with the highest error metrics, featuring a MAE of 8.89 and

an R^2 value of 0.4 compared to all other models. However, it's noteworthy that its performance surpasses that observed with the proximate and process parameters. Conversely, the Gradient Boosting and Random Forest models demonstrated superior performance within this dataset. Specifically, the Gradient Boosting model achieved a remarkable MAE of 4.82 with an outstanding R^2 of 0.85, indicating its high accuracy. The Random Forest model, while having a slightly higher MAE at 6.68, still displayed a commendable R^2 of 0.79. These results highlight the exceptional predictive capabilities of the Gradient Boosting model, making it the preferred choice for this dataset when utilizing ultimate analysis and process parameters.

Table 4.2.Error value for yield (%) prediction using ultimate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	7.31	8.89	6.68	4.82	8.8	8.32
MSE	85.83	139.02	68.5	41.36	117.54	96.24
RMSE	9.26	11.79	8.28	6.43	10.84	9.81
R^2	0.51	0.4	0.79	0.85	0.45	0.68

Following a comprehensive model comparison for both segments, the Gradient Boosting model emerged as the superior choice when utilizing ultimate analysis and process parameters. Subsequent hyperparameter tuning further improved its performance, resulting in a reduced error, with a MAE of 6.33 and an impressive R^2 value of 0.87. In Figure 4.5, the analysis of feature importance for yield (%) prediction is depicted. Notably, Ash (%) was identified as the most critical feature for accurate predictions, underscoring its pivotal role in the predictive model's performance.
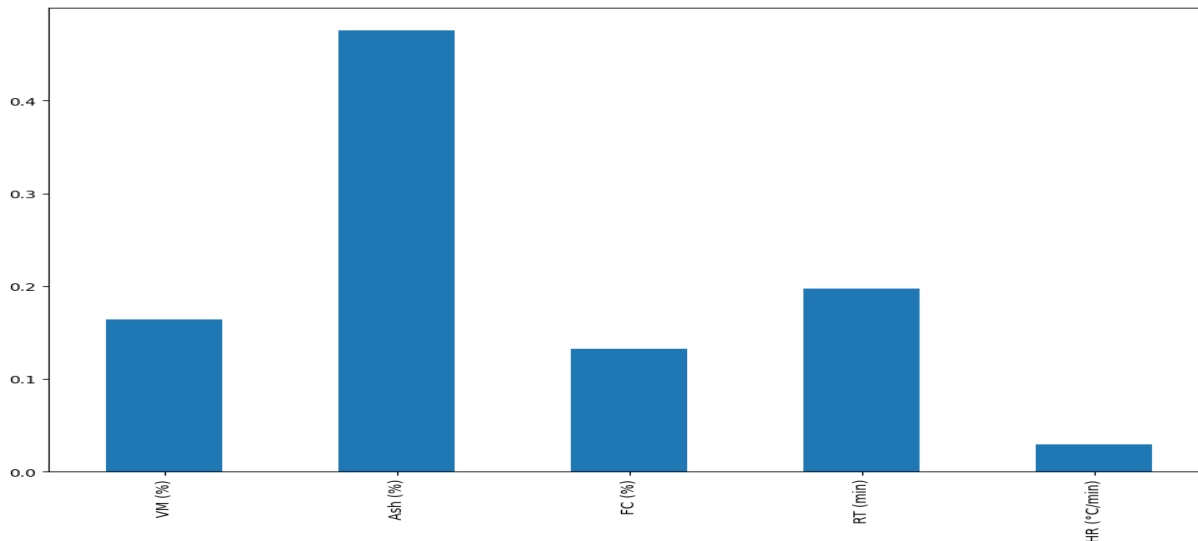
Figure 4.5 Feature Importance plot for Yield (%) prediction

4.2.2.2 Prediction of SSA

Utilizing the training dataset, predictive models were developed to estimate SSA, considering ultimate analysis, proximate analysis, and process parameters. In Figure 4.6, the performance of these models was assessed using a test dataset based on proximate analysis and process parameters. Notably, the Random Forest models and Adaptive Boosting models exhibited suboptimal predictive performance, as indicated by their respective predicted lines. Surprisingly, the Support Vector Regressor emerged as the most effective model during this dataset evaluation, with its predicted line closely resembling the ideal scenario where predicted values align with actual values (Y=X). This underscores the crucial importance of selecting the right algorithm, as it significantly influences prediction quality. In this context, the Support Vector Regressor models have demonstrated their potential to offer precise and reliable SSA predictions based on the specified parameters, indicating promising prospects for further analysis and real-world application.
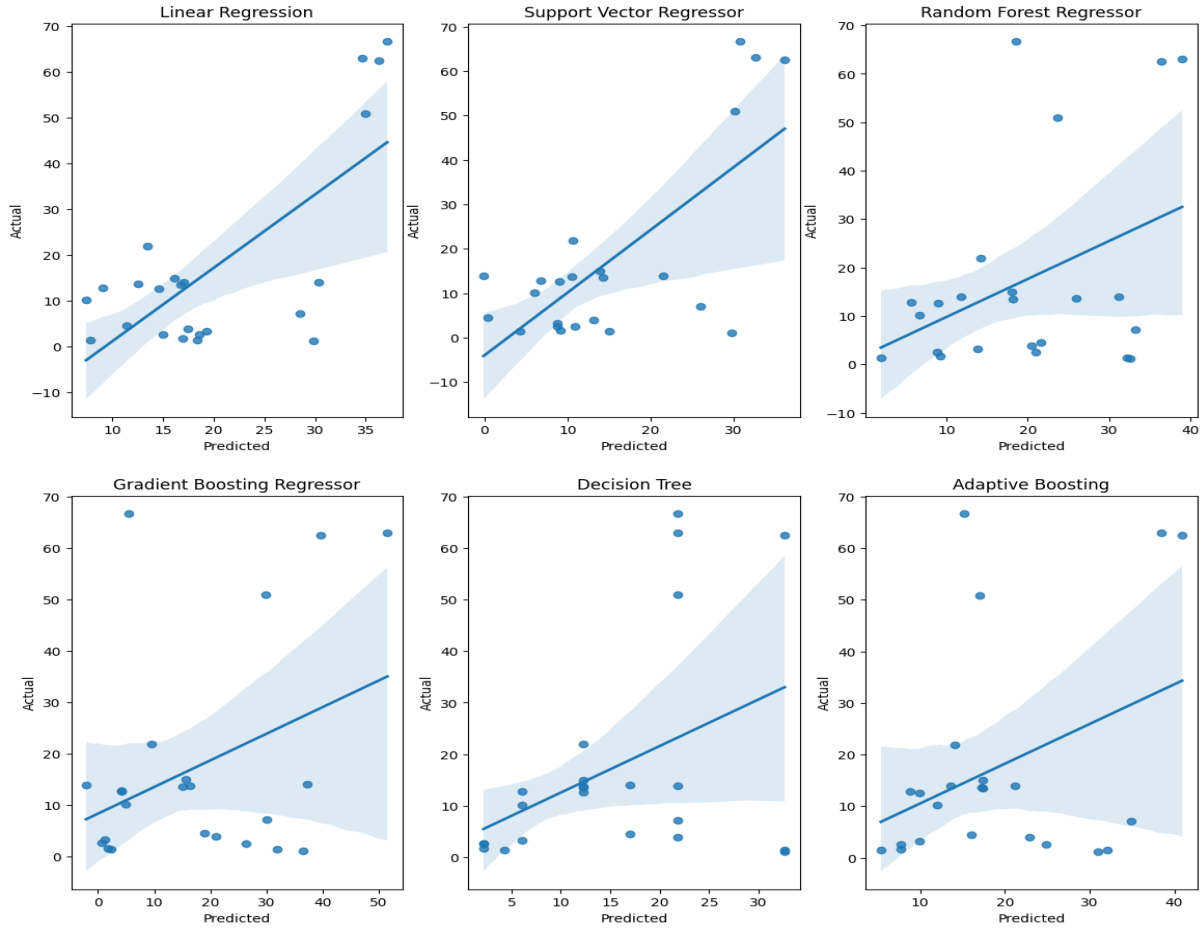
Figure 4.6 Regression plot for SSA using proximate and process parameters

Table 4.3 furnishes a comprehensive depiction of model performance, evaluated through a spectrum of pertinent metrics. Strikingly, the Random Forest models and Adaptive Boosting models surface with the most glaringly adverse error metrics. The Support Vector Regressor model exhibits a towering MAE of 54.81 and a conspicuously R^2 value of 0.07, while the Adaptive Boosting model features a prominent MAE of 50.79 and a notably negative R^2 of 0.63 in contrast to other models. Conversely, the Random Forest model emerges as the standout performer within this dataset, characterized by its exceptional predictive accuracy. Notably, it attains a significantly lower MAE of 22.58 and achieves a relatively positive R^2 of 0.88, underscoring its robust performance. This thorough analysis unequivocally establishes that when employing proximate analysis and process parameters, the Random Forest model offers the least prediction error with an MAE of 22.58 and the highest R^2 value of 0.88. These results accentuate the model's effectiveness in furnishing accurate SSA predictions grounded in the specified input parameters.

Table 4.3. Error value for SSA prediction using proximate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	72.33	54.81	22.58	27.67	36.0	50.59
MSE	7867.32	8337.9	1031.47	1472.68	2165.39	3323.88
RMSE	88.74	91.3	32.11	38.37	46.53	57.65
R^2	0.12	0.07	0.88	0.83	0.757	0.63

In Figure 4.7, the performance of machine learning models on a test dataset, employing ultimate analysis and process parameters, is illustrated. Impressively, all models showcased commendable performance in this dataset. It is important to note that while the Decision Tree and Adaptive Boosting models did not perform as strongly as some other models, they exhibited improved predictive capabilities compared to their performance with proximate and process parameters. The standout performers in this evaluation were the Support Vector Regressor models. These models excelled, with a substantial number of data points closely aligning with the ideal line, underscoring their effectiveness and precision in predicting the target variable within this specific dataset.
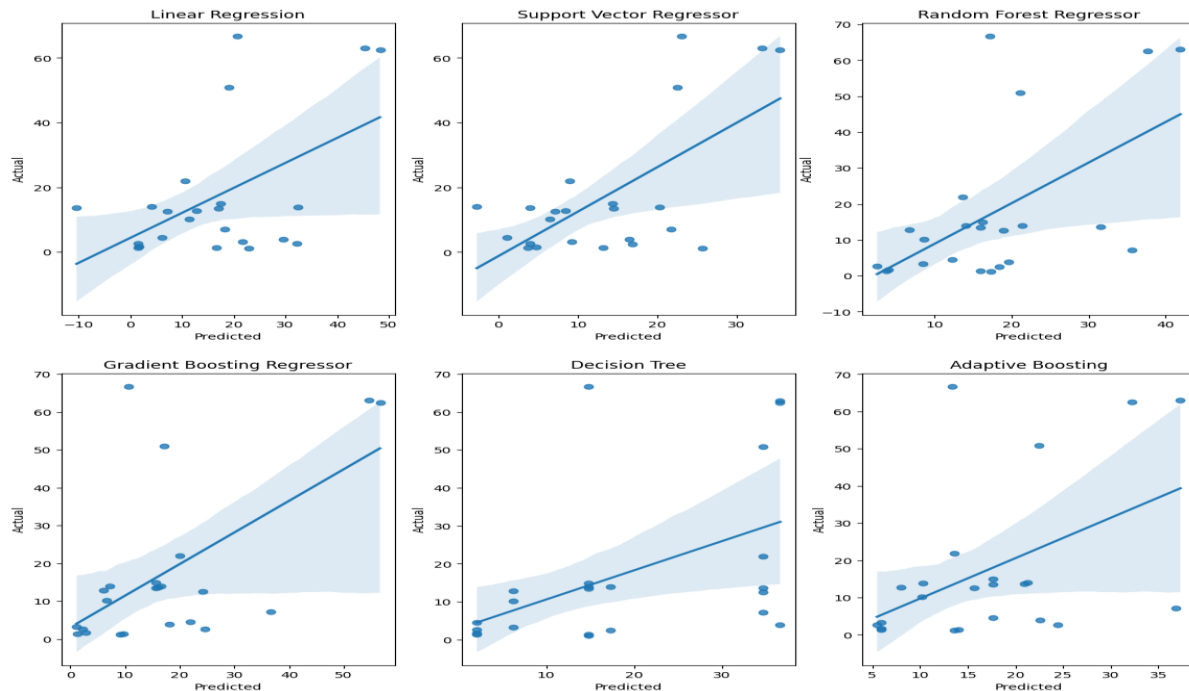

Figure 4.7 Regression plot for SSA using ultimate and process parameters

Table 4.4 offers a comprehensive analysis of model performance, assessed through various evaluation metrics. Notably, the Support Vector Regressor model stands out with the highest error metrics, featuring a considerable MAE of 53.43 and a R^2 value of 0.027. However, it's essential to recognize that their performance outperforms what was observed with proximate and process parameters. Conversely, the random models exhibited superior performance within this dataset. Specifically, the model achieved an impressive MAE of 26.93 with a remarkable R^2 of 0.83, underscoring its high accuracy and predictive capabilities. These outcomes underscore the exceptional performance of the random forest model, positioning it as the preferred choice for this dataset when utilizing ultimate analysis and process parameters.

Table 4.4.Error value for SSA prediction using ultimate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	76.33	53.43	26.93	34.78	52.87	47.32
MSE	8277.59	8669.71	1465.05	2305.95	5689.46	3290.8
RMSE	90.98	93.10	38.27	48.02	75.42	57.36
R^2	0.071	0.027	0.83	0.74	0.36	0.63

After conducting a thorough model comparison for both data segments, the random forest model emerged as the optimal choice when employing proximate analysis and process parameters. Subsequent fine-tuning of its hyperparameters further enhanced its performance, resulting in a notable reduction in prediction errors. The random forest model achieved an impressive MAE of 22.58, coupled with an outstanding R^2 value of 0.88, highlighting its superior accuracy in predicting the target variable.

In Figure 4.8, an analysis of feature importance for predicting SSA is presented. Notably, two features, FC (%) and Ash (%), were identified as the most critical contributors to accurate predictions. This underscores the pivotal role played by these features in driving the overall performance of the predictive model, emphasizing their significance in achieving precise yield predictions.
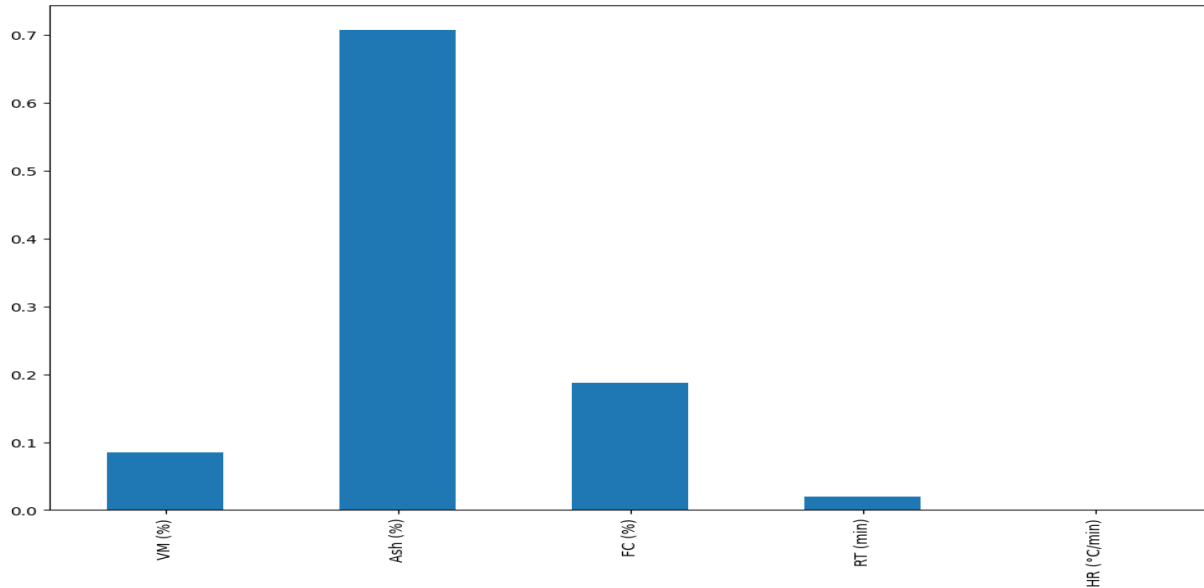
Figure 4.8 Feature Importance plot for SSA prediction

4.2.2.3 Prediction of HHV

Using the training dataset, predictive models were constructed to estimate HHV, taking into account ultimate analysis, proximate analysis, and process parameters. In Figure 3.9, the performance of these models was evaluated using a test dataset that relied on proximate analysis and process parameters. Notably, the Linear Regression model displayed suboptimal predictive performance, as evidenced by its predicted lines.

Surprisingly, the Random Forest and Adaptive Boosting models emerged as the most effective options during this dataset evaluation, with their predicted lines closely resembling the ideal scenario where predicted values align perfectly with actual values (Y=X). This highlights the pivotal role of algorithm selection, which significantly impacts the quality of predictions. In this context, the Adaptive Boosting models have proven their potential to provide precise and reliable HHV predictions based on the specified parameters. This finding suggests promising avenues for further analysis and practical applications, underscoring the importance of choosing the right model for the task at hand.
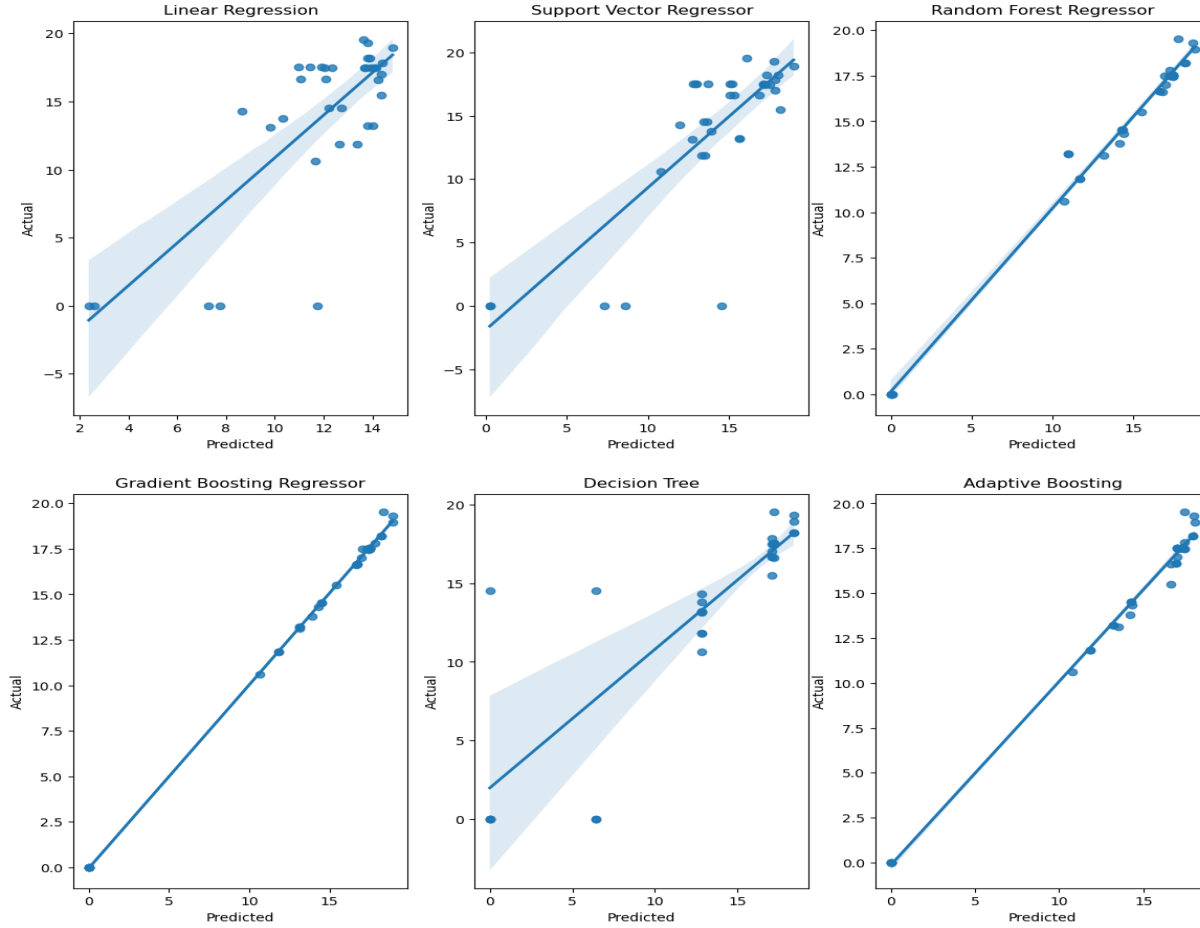
Figure 4.9 Regression plot for HHV using Proximate and process parameters

Table 4.5 provides a comprehensive overview of model performance, assessed through a range of key metrics. The most striking observation is that the Linear Regression model exhibits notably unfavourable error metrics. Specifically, it displays a substantial MAE of 3.99 and a R^2 value of 0.83 compared to the other models. Conversely, the Gradient Boosting model emerges as the standout performer within this dataset, marked by its exceptional predictive accuracy. Impressively, it achieves a substantially lower MAE of 0.08 and a notably positive R^2 of 0.96, underscoring its robust and highly accurate performance. This comprehensive analysis unequivocally establishes that when utilizing proximate analysis and process parameters, the Gradient Boosting model excels by offering the least prediction error. These findings emphasize the model's effectiveness in providing accurate predictions for HHV, based on the specified input parameters.

Table 4.5.Error value for HHV prediction using proximate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	3.99	2.22	0.3	0.08	1.51	0.29
MSE	21.15	13.35	0.41	0.05	10.57	0.25
RMSE	4.6	3.65	0.64	0.22	3.25	0.5
R^2	0.83	0.87	0.91	0.96	0.69	0.78

Figure 4.10 presents the performance of machine learning models on a test dataset, leveraging ultimate analysis and process parameters. Impressively, all models demonstrated commendable performance in this dataset. Notably, while Linear Regression didn't perform as strongly as some other models, it doesn't exhibit improved predictive capabilities compared to its performance with proximate and process parameters. The standout performers in this evaluation were the Random Forest and Gradient Boosting models. These models excelled, with a substantial number of data points closely aligning with the ideal line, underscoring their effectiveness and precision in predicting the target variable within this specific dataset. This emphasizes the critical role of model selection, as these models proved their worth by delivering highly accurate predictions. Ultimately, the dataset, incorporating ultimate analysis and process parameters, benefited significantly from the Random Forest and Gradient Boosting models, which offer promising avenues for advanced analysis and real-world applications.
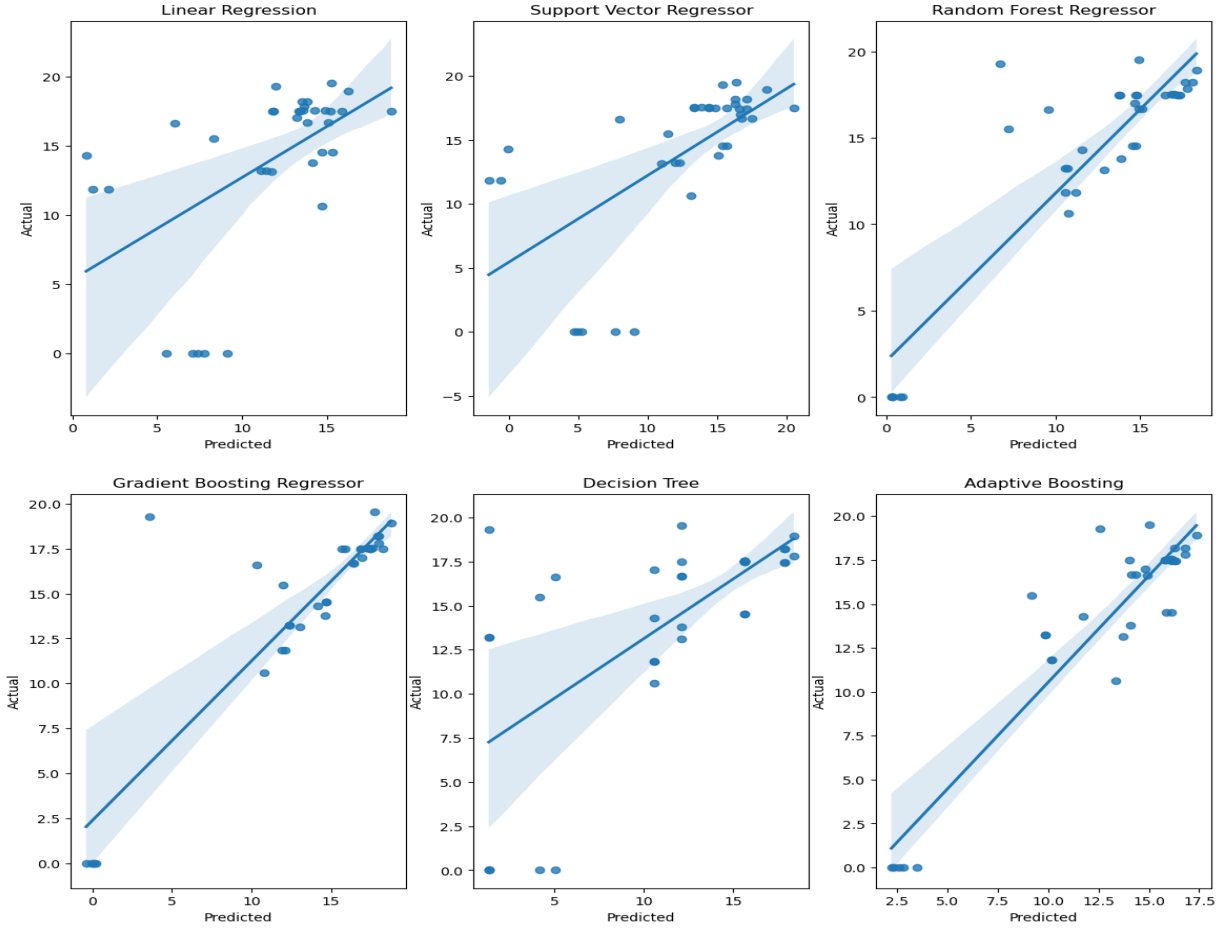
**Figure 4.10 Regression plot for HHV using ultimate and process parameters**

Table 4.6 provides a comprehensive analysis of model performance, assessed through various evaluation metrics. Notably, the Linear Regression model stands out with the highest error metrics, featuring a considerable MAE of 4.71 and a R^2 value of 0.61, when compared to the other models. However, it's crucial to acknowledge that its performance surpasses what was observed with proximate and process parameters.

Conversely, the random forest model demonstrated superior performance within this dataset. Specifically, the model achieved an impressive MAE of 1.92 with a remarkable R^2 of 0.95, underscoring its high accuracy and predictive capabilities. These results underscore the exceptional performance of the Gradient Boosting model, positioning it as the preferred choice for this dataset when utilizing ultimate analysis and process parameters. This signifies the model's effectiveness in providing accurate predictions, highlighting its potential for advanced analysis and practical applications.

Table 4.6.Error value for HHV prediction using ultimate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	4.71	3.74	1.92	1.06	3.7	1.98
MSE	32.26	27.21	10.56	8.38	31.5	9.0
RMSE	5.68	5.22	3.25	2.9	5.61	3.0
R^2	0.61	0.64	0.95	0.94	0.73	0.91

Following an extensive model comparison for both data segments, the Gradient Boosting model emerged as the optimal choice when utilizing proximate analysis and process parameters. The random forest model achieved exceptional results, with a strikingly low MAE of 0.08 and an impressive R^2 value of 0.96, underscoring its outstanding accuracy in predicting the target variable. Figure 3.11 delves into the analysis of feature importance for predicting HHV. Notably, two key features, VM (%), and FC (%), were identified as the most critical contributors to accurate predictions. This underscores the pivotal role played by these features in driving the overall performance of the predictive model, underscoring their significance in achieving highly precise yield predictions. These findings highlight the potential of the Gradient Boosting model and the pivotal role of these specific features in enhancing the accuracy and reliability of HHV predictions, offering promising prospects for further analysis and practical application.
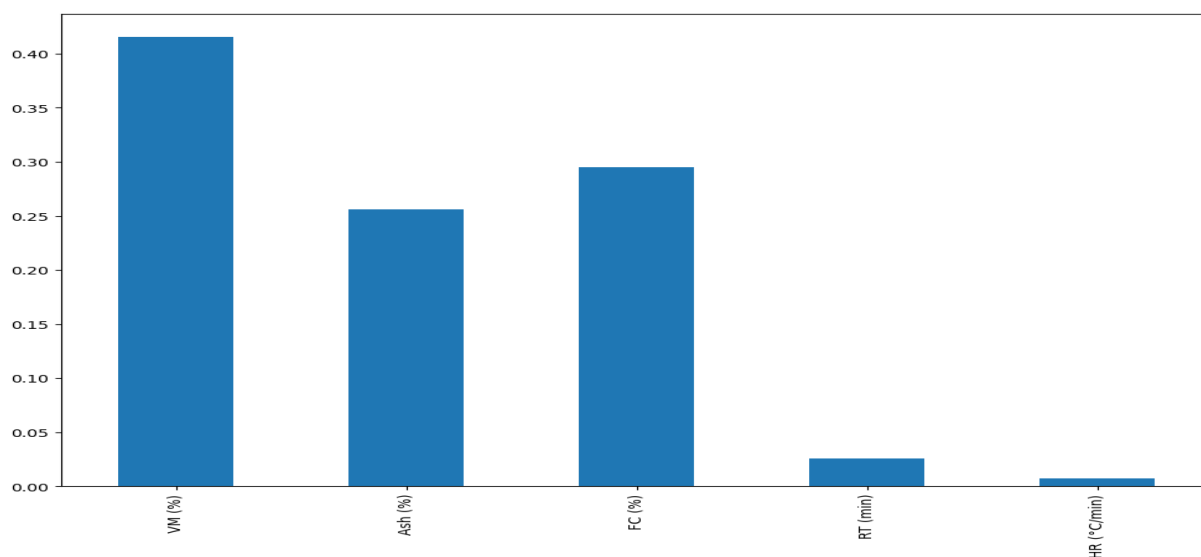
Figure 4.11 Feature Importance plot for HHV prediction

4.3 Modelling for Biooil

The prediction goal is divided into three sections, aligning with each target variable (yield, Vis, HHV). Furthermore, within each section, two techniques are utilized: (a) utilizing proximate analysis and process parameters, and (b) employing ultimate analysis and process parameters. Consequently, there are a total of 12 distinct models, each crafted to predict one of the three properties based on the chosen parameter set.

4.3.1 Effect of Process Parameters

Figure 4.12 and Table 2.2 provide visual representations of our training data, presenting scatter plots that illustrate the relationships between different features and their effects on yield (%), Vis, and HHV. The scatter plots are meticulously arranged to demonstrate how variations in each feature correspond to these crucial variables. Notably, the histograms displayed along the diagonal of the scatter plots offer significant insights. These histograms reveal that the distribution of our features closely resembles a normal distribution, a vital assumption in numerous statistical analyses and machine learning algorithms, thus bolstering the reliability of our findings.

Upon examining the relationships between yield (%) and our features, it becomes evident that C (%), H (%), PS, and T(°C) exhibit a positive correlation with yield (%). Conversely, viscosity shows a positive association with N (%) and FC (%). Furthermore, HHV demonstrates a positive relationship with FC, T(°C), and HR (°C/min). The yield histogram highlights that the central tendency of all experimentally determined yields centres around 40%, offering pivotal insights into the overall performance and characteristics of our dataset. This statistic serves as a crucial reference point for subsequent analysis and modelling endeavours.
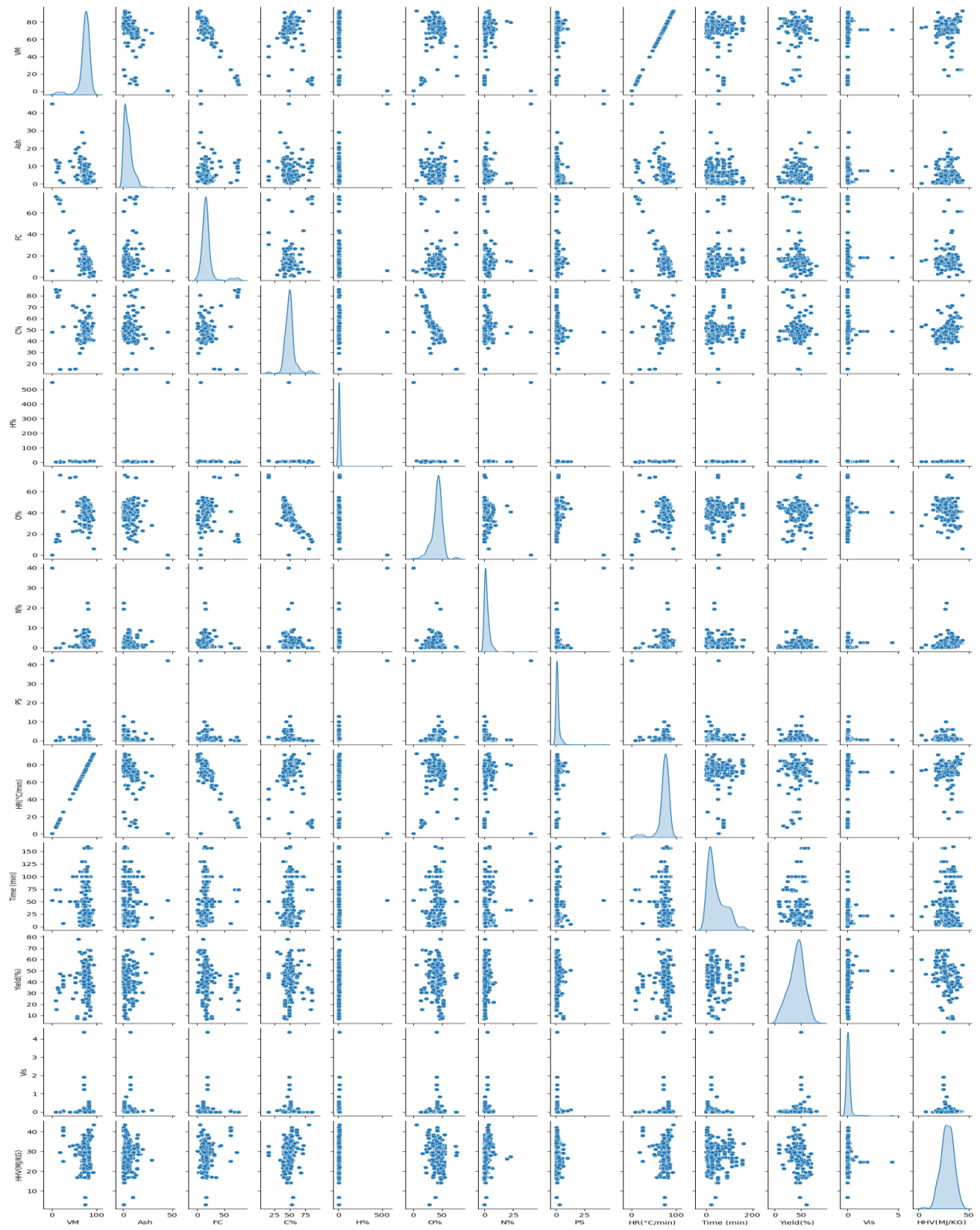
Figure 4.12 Scatter plot of the various process parameters and Target Variable

Figure 4.13 explores the intricate relationship between different process parameters and the target variable. A notable finding is the minimal correlation among the process parameters, indicating the absence of multicollinearity. Multicollinearity, if present, can weaken the statistical power of regression analyses, making its absence in this context a favourable characteristic. Yield (%) demonstrates strong positive correlations with H (%), T (°C), and RT (min), with correlation coefficients of 0.32, 0.21, and 0.15, respectively. Conversely, it shows a significant negative correlation with O (%), with a correlation coefficient of -0.21. Viscosity exhibits robust positive correlations with N (%), boasting correlation coefficients of 0.14, and noteworthy negative correlations with FC (%), with correlation coefficients of -0.13. In its relationship with the process parameters, HHV shows significant positive correlations with O (%), N (%), and FC (%), yielding correlation coefficients of 0.16, 0.18, and 0.31, respectively. Conversely, it showcases substantial negative correlations with VM (%) and C (%), featuring correlation coefficients of -0.37 and -0.12, respectively. Despite these correlations, the scatter plots in Figure 4.13 lack a discernible trend, suggesting that a comprehensive experimental analysis of various process parameter combinations may be necessary. This highlights the crucial role of advanced machine learning algorithms in constructing predictive models capable of quantifying yield as a function of these multifaceted process parameters.
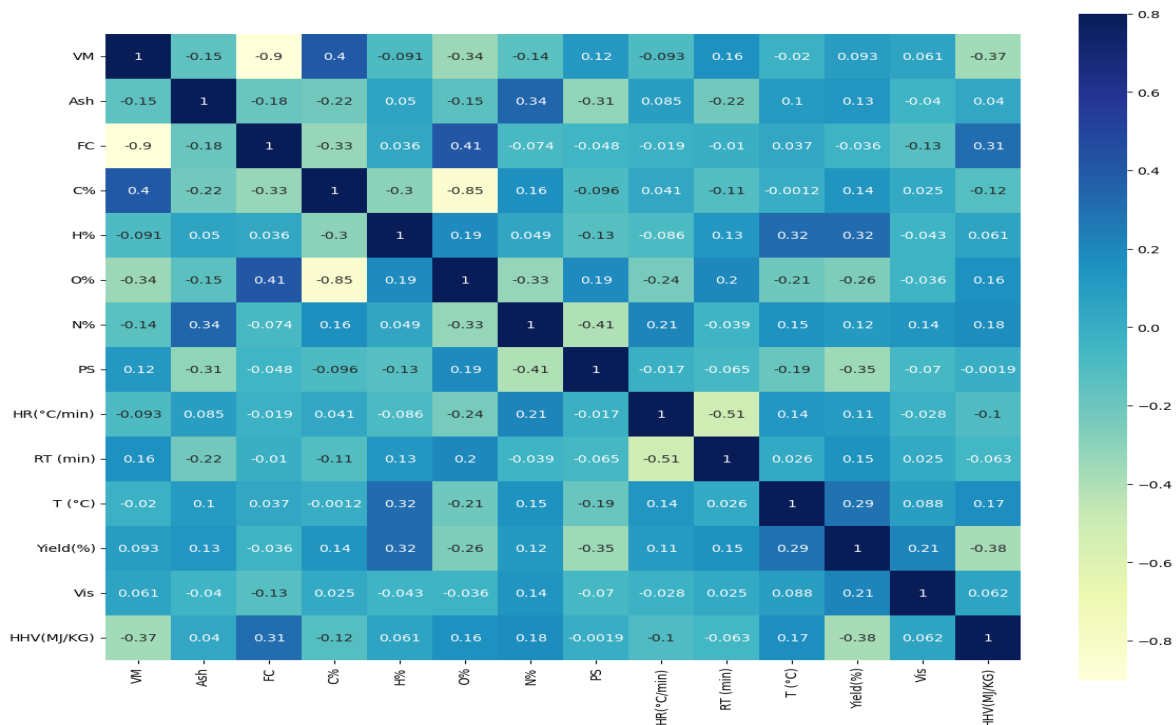

Figure 4.13 Correlation between the various process parameters and Target Variable

4.3.2 Comparison of Various ML Predictive Models

The evaluation and comparison of machine learning predictive models will occur separately for each of the three distinct segments. Regression plots (regplots) will visually demonstrate the relationships between actual and predicted values. Additionally, comprehensive model evaluation will encompass key metrics such as MAE, MSE, RMSE and the coefficient of determination R^2. Model selection will prioritize minimizing prediction errors (MAE, MSE, RMSE) and maximizing goodness-of-fit (R^2). The selected model, with the least error and highest R^2, will undergo hyperparameter tuning to enhance predictive accuracy further. Subsequently, the final optimized model will undergo feature importance analysis, visually represented by a bar graph. This analysis will elucidate the relative significance of each feature, aiding in model interpretability and feature selection.

4.3.2.1 Prediction of Yield (%)

Using the training dataset, predictive models were developed to predict the yield percentage (%), incorporating ultimate analysis, proximate analysis, and process parameters. Figure 4.14 illustrates the performance of machine learning models on a test dataset, utilizing proximate analysis and process parameters. Remarkably, all models demonstrated impressive performance on this dataset. Particularly noteworthy were the Support Vector Regressor models, which stood out in this assessment. These models performed exceptionally well, with a substantial proportion of data points closely matching the ideal Y=X line, underscoring their efficacy and precision in predicting the target variable within this dataset.
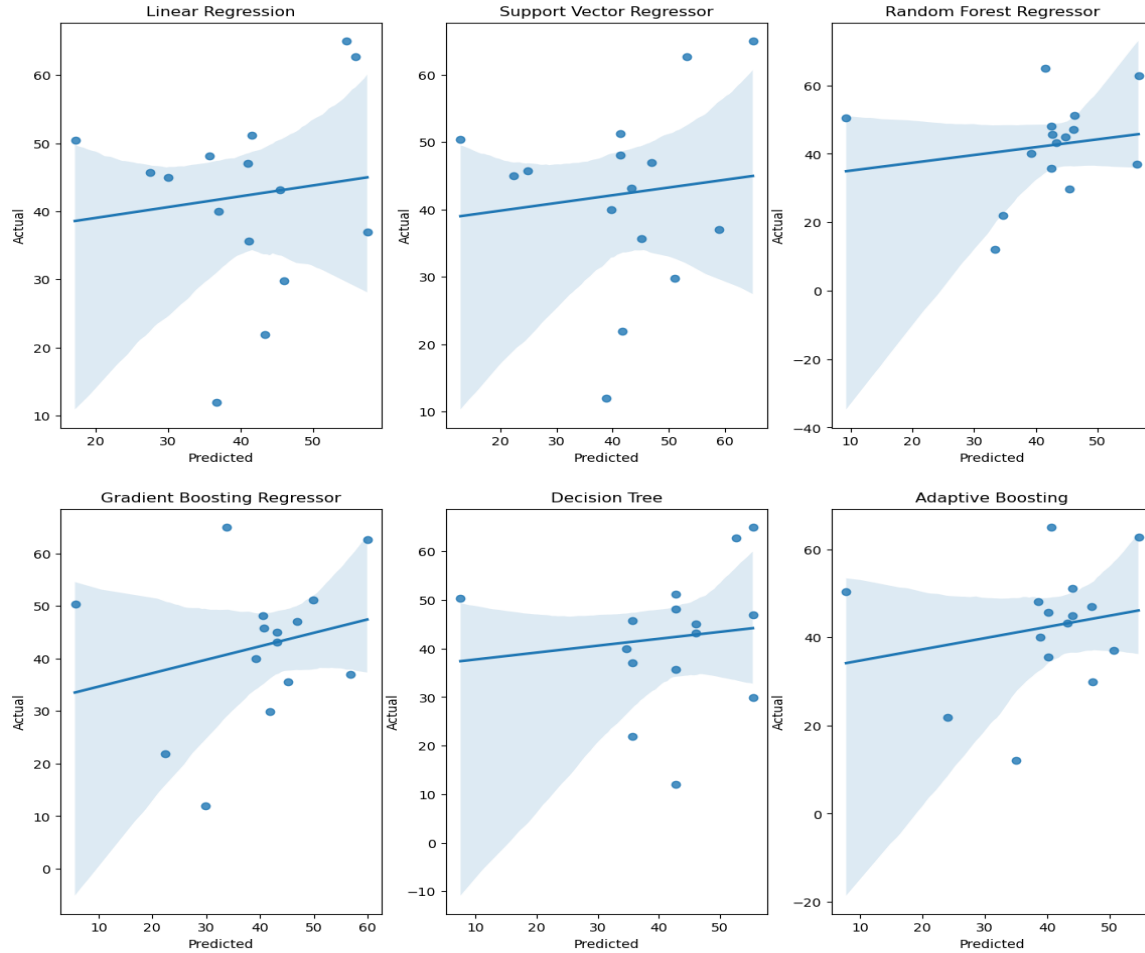
Figure 4.14 Regression plot for Yield (%) using proximate and process parameters

Table 4.7 provides a thorough examination of model performance using evaluation metrics. Notably, its performance exceeds that observed with the proximate and process parameters. Conversely, the Support Vector Regressor model exhibited superior performance within this dataset. Particularly, the Support Vector Regressor model achieved an impressive MAE of 10.32 with an outstanding R^2 of 0.63. These findings underscore the exceptional predictive prowess of the Support Vector Regressor model, establishing it as the preferred option for this dataset when employing ultimate analysis and process parameters.

Table 4.7.Eror value for yield (%) prediction using proximate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	12.51	10.32	11.38	14.63	13.81	12.62
MSE	243.43	196.42	251.5	364.36	323.57	296.64
RMSE	15.6	14.01	15.85	19.08	17.98	17.22
R^2	0.45	0.63	0.32	0.33	0.39	0.31

Figure 4.15 assesses the performance of predictive models using a test dataset reliant on ultimate analysis and process parameters. Surprisingly, the Support Vector Regressor exhibited the least favourable predictive performance, evident from its predicted line. Conversely, the Decision Tree and adaptive boosting models emerged as the most effective within this dataset evaluation. Their predicted lines closely resemble the ideal scenario where predicted values match actual values (Y=X), showcasing their superior performance and predictive accuracy. This outcome underscores the importance of model selection, as the choice of an appropriate algorithm significantly influences prediction quality. In this context, Decision Tree and adaptive boosting models have demonstrated potential to provide precise and reliable yield predictions concerning specified parameters, presenting promising avenues for further analysis and application.
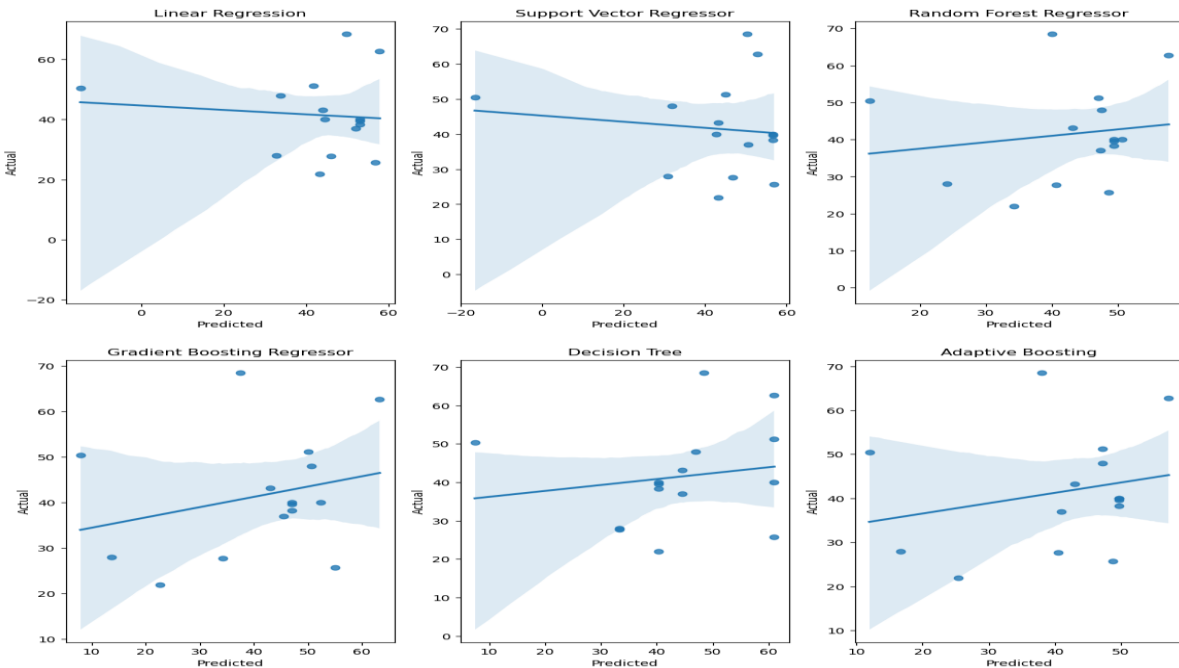

Figure 4.15 Regression plot for Yield (%) using ultimate and process parameters

Table 4.8 presents a thorough examination of model performance using various evaluation metrics. Notably, the Support Vector Regressor stands out with the highest error metrics, showcasing a MAE of 13.39 and a notably R^2 of 0.14 compared to all other models. In contrast, the Decision Tree and adaptive boosting models emerge as the top-performing choices within this dataset, distinguished by their exceptional predictive accuracy. The Decision Tree model reports an MAE of 10.65, along with an R^2 of 0.54, demonstrating its robust performance. Meanwhile, the adaptive boosting model achieves an even lower MAE of 10.5 and an impressive R^2 of 0.58.

This analysis conclusively establishes that when utilizing ultimate analysis and process parameters, the adaptive boosting model emerges as the model with the least prediction error MAE of 10.5 and the highest R^2 value 0.58. These findings underscore the effectiveness of this model in delivering precise yield predictions based on the specified input parameters.

Table 4.8.Error value for yield (%) prediction using ultimate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	13.09	13.39	12.23	10.33	10.65	10.5
MSE	259.23	314.49	242.52	268.13	279.8	246.69
RMSE	16.10	17.73	15.59	16.38	16.72	15.7
R^2	0.34	0.14	0.45	0.52	0.54	0.58

After conducting a thorough comparison of models for both segments, the Support Vector regressor model emerged as the optimal option when utilizing proximate analysis and process parameters. Further hyperparameter tuning significantly enhanced its performance, leading to decreased error, with a MAE of 10.32 and an impressive R^2 value of 0.64. Figure 4.5 illustrates the analysis of feature importance for predicting yield (%). Notably, PS was identified as the most crucial feature for accurate predictions, highlighting its pivotal role in enhancing the predictive model's performance.
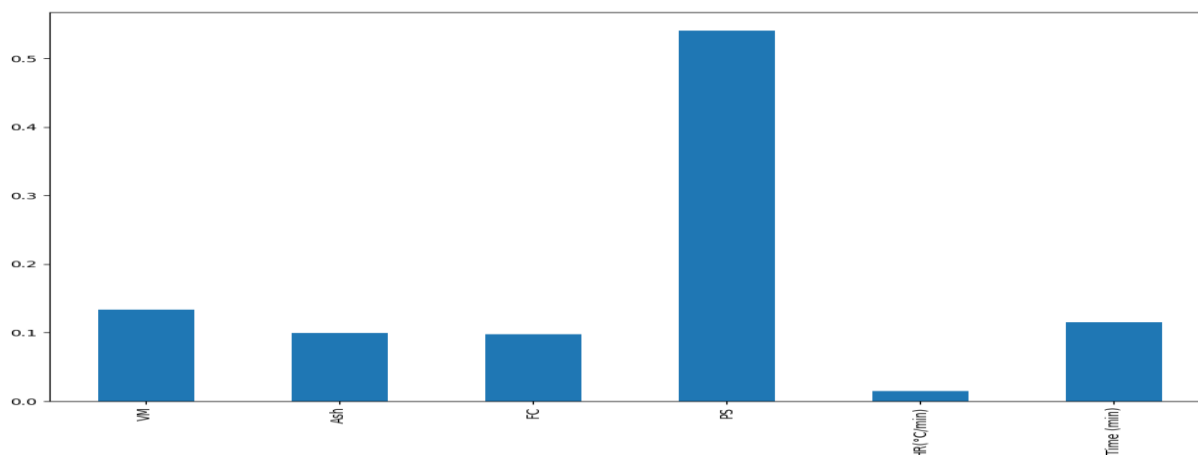
Figure 4.16 Feature Importance plot for Yield (%) prediction

4.3.2.2 Prediction of Viscosity

Using the training dataset, predictive models were developed to estimate Viscosity, incorporating ultimate analysis, proximate analysis, and process parameters. In Figure 4.17, the performance of these models was evaluated using a test dataset based on proximate analysis and process parameters. Surprisingly, the Decision Tree emerged as the most effective model during this dataset evaluation, closely resembling the ideal scenario where predicted values align with actual values (Y=X). This emphasizes the critical importance of selecting the appropriate algorithm, as it significantly impacts prediction quality. In this context, the Decision Tree model has demonstrated its potential to provide precise and reliable Viscosity predictions based on the specified parameters, suggesting promising avenues for further analysis and real-world application.
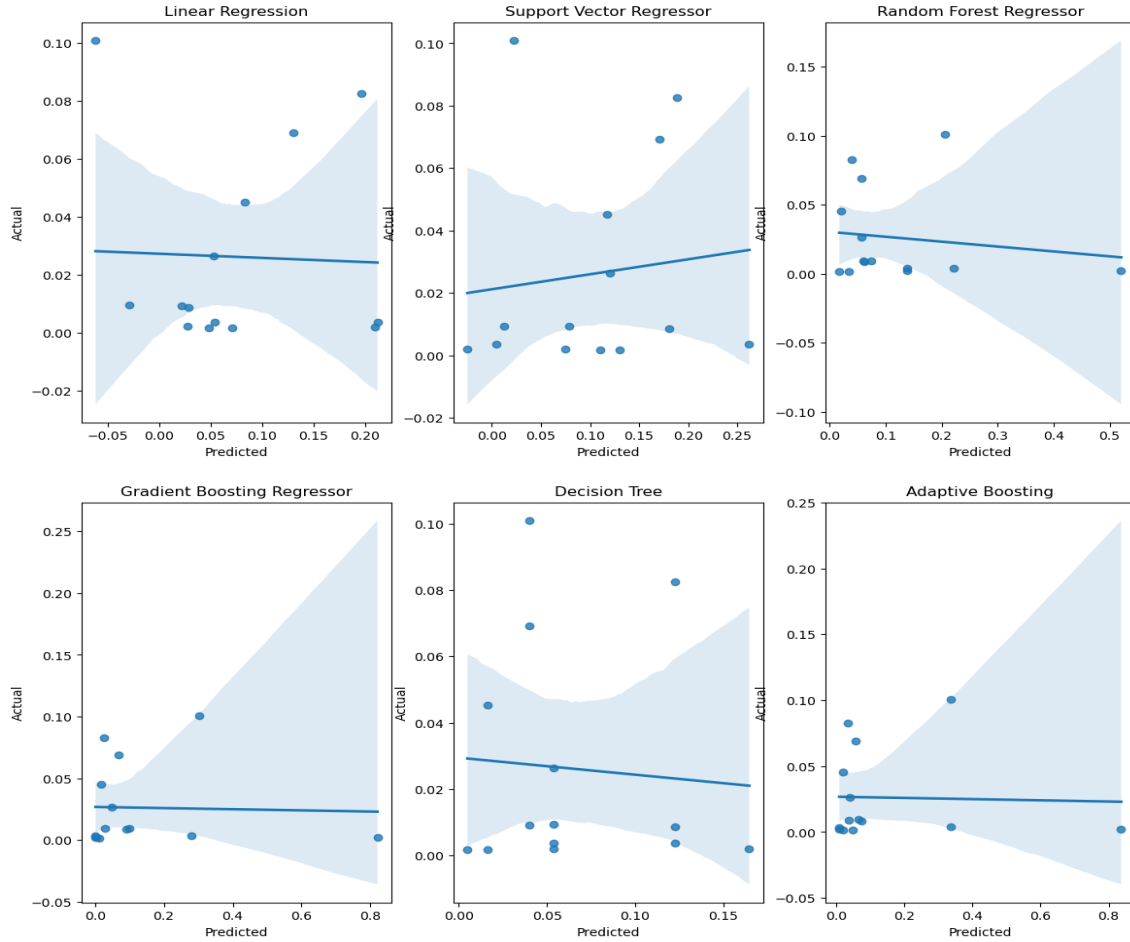
Figure 4.17 Regression plot for Viscosity using proximate and process parameters

Table 4.9 provides a comprehensive overview of model performance, assessed through a range of relevant metrics. The Decision Tree Regressor model stands out as the top performer within this dataset, distinguished by its exceptional predictive accuracy. Notably, it achieves a substantially lower MAE of 0.05 and obtains a relatively positive R^2 of 0.49, highlighting its strong performance.

Table 4.9.Error value for Viscosity prediction using proximate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	0.08	0.09	0.11	0.12	0.05	0.12
MSE	0.01	0.13	0.02	0.06	0.01	0.06
RMSE	0.39	0.66	0.48	0.56	0.39	0.25
R^2	0.24	0.41	0.34	0.35	0.49	0.31

Figure 4.18 depicts the performance of machine learning models on a test dataset utilizing ultimate analysis and process parameters. Remarkably, all models demonstrated commendable performance in this dataset. The standout performers in this evaluation were the Decision Tree models. These models excelled, with a significant number of data points closely aligning with the ideal line, emphasizing their effectiveness and precision in predicting the target variable within this specific dataset.
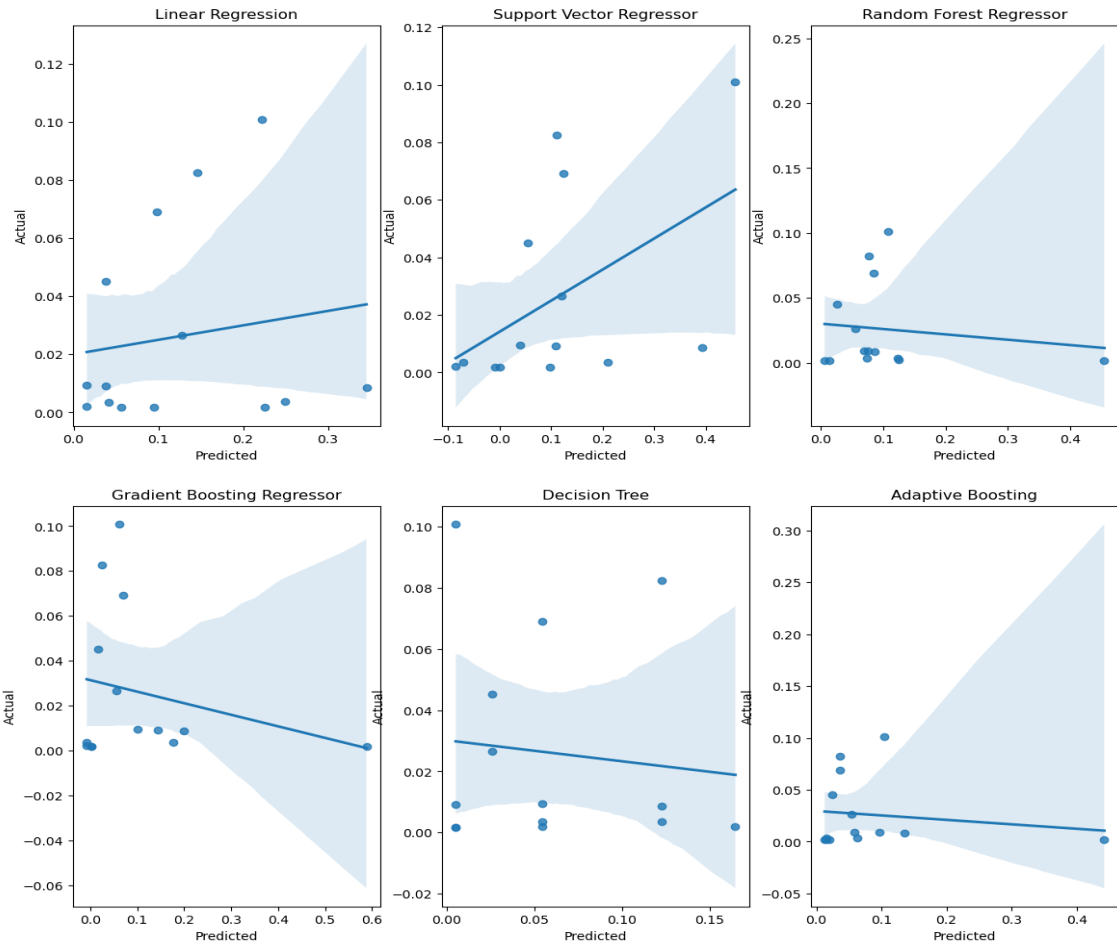

Figure 4.18 Regression plot for SSA using ultimate and process parameters

Table 4.10 provides an in-depth analysis of model performance, evaluated using various assessment metrics. Importantly, it's crucial to acknowledge that their performance surpasses what was observed with proximate and process parameters. In contrast, the Decision Tree models demonstrated superior performance within this dataset. Specifically, the model achieved an impressive MAE of 0.01 with a remarkable R^2 of 0.51, highlighting its high accuracy and predictive capabilities. These results underscore the exceptional performance of the Decision

Tree model, establishing it as the preferred option for this dataset when utilizing ultimate analysis and process parameters.

Table 4.10.Error value for Viscosity prediction using ultimate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	0.09	0.11	0.08	0.010	0.01	0.07
MSE	0.02	0.03	0.02	0.03	0.005	0.01
RMSE	18.13	16.63	17.21	16.58	18.55	16.43
R^2	0.23	0.44	0.31	0.40	0.51	0.34

Following a comprehensive model comparison across both data segments, the Decision Tree model emerged as the optimal choice when utilizing ultimate analysis and process parameters. Subsequent fine-tuning of its hyperparameters further bolstered its performance, resulting in a notable reduction in prediction errors. The Decision Tree model achieved an impressive MAE of 0.01, accompanied by an outstanding R^2 value of 0.51, highlighting its superior accuracy in predicting the target variable. Figure 4.19 presents an analysis of feature importance for predicting Viscosity. Remarkably, two features, Ash (%) and Time, were identified as the most critical contributors to accurate predictions. This emphasizes the pivotal role played by these features in driving the overall performance of the predictive model, underscoring their significance in achieving precise Viscosity predictions.
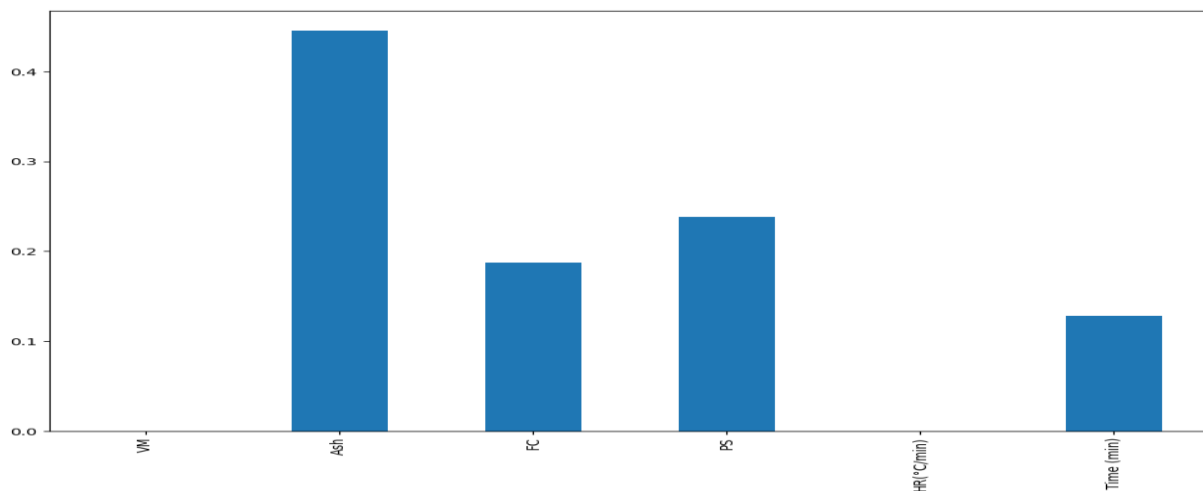

Figure 4.19 Feature Importance plot for Viscosity prediction

4.3.2.3 Prediction of HHV

Using the training dataset, predictive models were developed to estimate HHV, incorporating ultimate analysis, proximate analysis, and process parameters. In Figure 4.20, the performance of these models was assessed using a test dataset based on proximate analysis and process parameters. Notably, the Linear Regression model demonstrated suboptimal predictive performance, as indicated by its predicted lines.

Surprisingly, the Random Forest and Adaptive Boosting models emerged as the most effective options during this dataset evaluation, with their predicted lines closely resembling the ideal scenario where predicted values perfectly align with actual values (Y=X). This underscores the critical role of algorithm selection, which significantly influences prediction quality. In this context, the Random Forest models have demonstrated their potential to offer precise and reliable HHV predictions based on the specified parameters. This discovery suggests promising opportunities for further analysis and practical applications, emphasizing the importance of selecting the appropriate model for the task at hand.
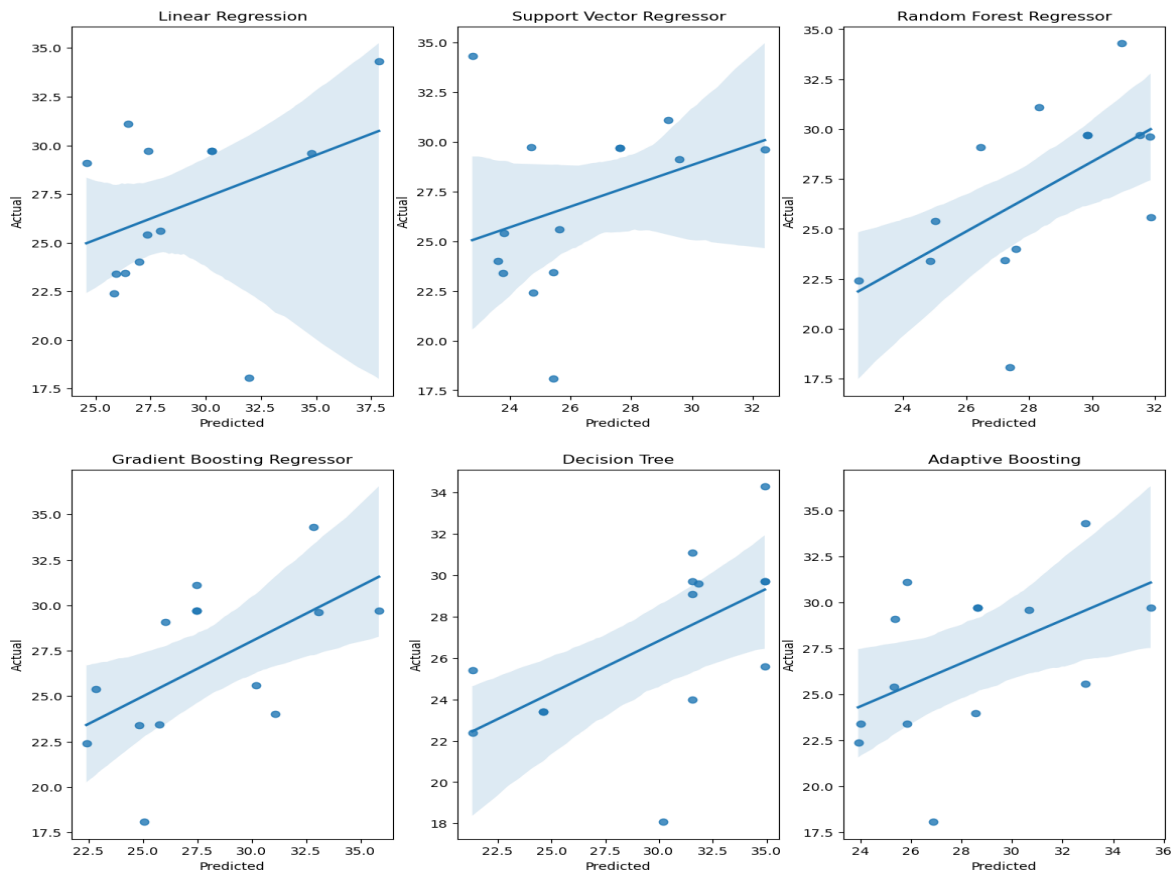

Figure 4.20 Regression plot for HHV using Proximate and process parameters

Table 4.11 offers a comprehensive summary of model performance, evaluated across various key metrics. The Random Forest model stands out as the top performer within this dataset, distinguished by its exceptional predictive accuracy. Impressively, it achieves a significantly lower MAE of 2.71 and a notably positive R^2 of 0.71, highlighting its robust and highly accurate performance. This thorough analysis unequivocally establishes that when employing proximate analysis and process parameters, the Random Forest model excels by minimizing prediction error. These findings underscore the model's effectiveness in delivering accurate predictions for HHV, based on the specified input parameters.

Table 4.11.Error value for HHV prediction using proximate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	3.66	2.85	2.71	3.86	3.89	3.14
MSE	23.18	17.45	13.48	15.5	27.17	17.21
RMSE	4.81	4.18	3.67	3.93	5.21	4.15
R^2	0.31	0.42	0.71	0.65	0.45	0.52

Figure 4.21 illustrates the performance of machine learning models on a test dataset utilizing ultimate analysis and process parameters. Impressively, all models demonstrated commendable performance in this dataset. While Linear Regression didn't perform as strongly as some other models, it did not exhibit improved predictive capabilities compared to its performance with proximate and process parameters. The standout performer in this evaluation was the Adaptive Boosting models. These models excelled, with a substantial number of data points closely aligning with the ideal line, underscoring their effectiveness and precision in predicting the target variable within this specific dataset. This highlights the critical role of model selection, as these models proved their worth by delivering highly accurate predictions. Ultimately, the dataset, incorporating ultimate analysis and process parameters, greatly benefited from the Adaptive Boosting models, which offer promising avenues for advanced analysis and real-world applications.
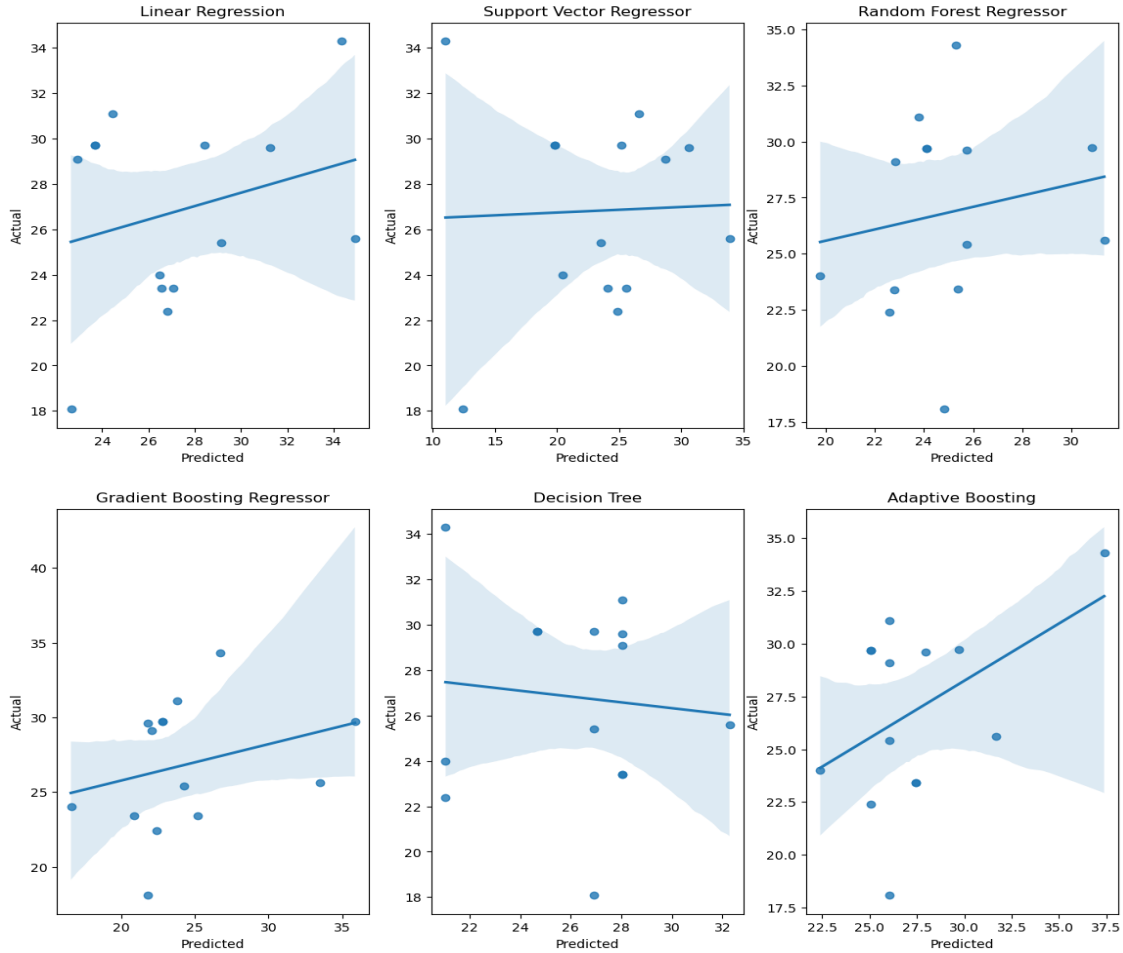
Figure 4.21 Regression plot for HHV using ultimate and process parameters

Table 4.12 offers an extensive assessment of model performance, evaluated across various evaluation metrics. The Adaptive Boosting models exhibited superior performance within this dataset. Specifically, the model achieved an impressive MAE of 3.51 with a remarkable R^2 of 0.65, emphasizing its high accuracy and predictive capabilities. These findings highlight the exceptional performance of the Adaptive Boosting model, establishing it as the preferred option for this dataset when utilizing ultimate analysis and process parameters. This underscores the model's effectiveness in delivering accurate predictions, indicating its potential for advanced analysis and practical applications.

Table 4.12. Error value for HHV prediction using ultimate and process parameters

	Linear Regression	SVR	Random Forest	Gradient Boosting	Decision Tree	AdaBoost
MAE	4.23	5.58	4.18	5.28	4.45	3.51
MSE	23.6	64.82	25.2	35.35	13.32	16.61
RMSE	4.85	8.05	5.02	5.94	5.51	4.08
R2	0.51	0.45	0.60	0.43	0.35	0.65

After an extensive comparison of models across both data segments, the Random Forest model emerged as the optimal choice when utilizing proximate analysis and process parameters. The Random Forest model yielded exceptional results, boasting a remarkably low MAE of 2.71 and an impressive R^2 value of 0.71, highlighting its outstanding accuracy in predicting the target variable. Figure 4.22 provides an in-depth analysis of feature importance for predicting HHV. Remarkably, three key features, VM (%), FC (%), and Time (min), were identified as the most critical contributors to accurate predictions. This underscores the pivotal role played by these features in driving the overall performance of the predictive model, emphasizing their significance in achieving highly precise yield predictions. These findings underscore the potential of the Gradient Boosting model and the crucial role of these specific features in enhancing the accuracy and reliability of HHV predictions, offering promising prospects for further analysis and practical application.
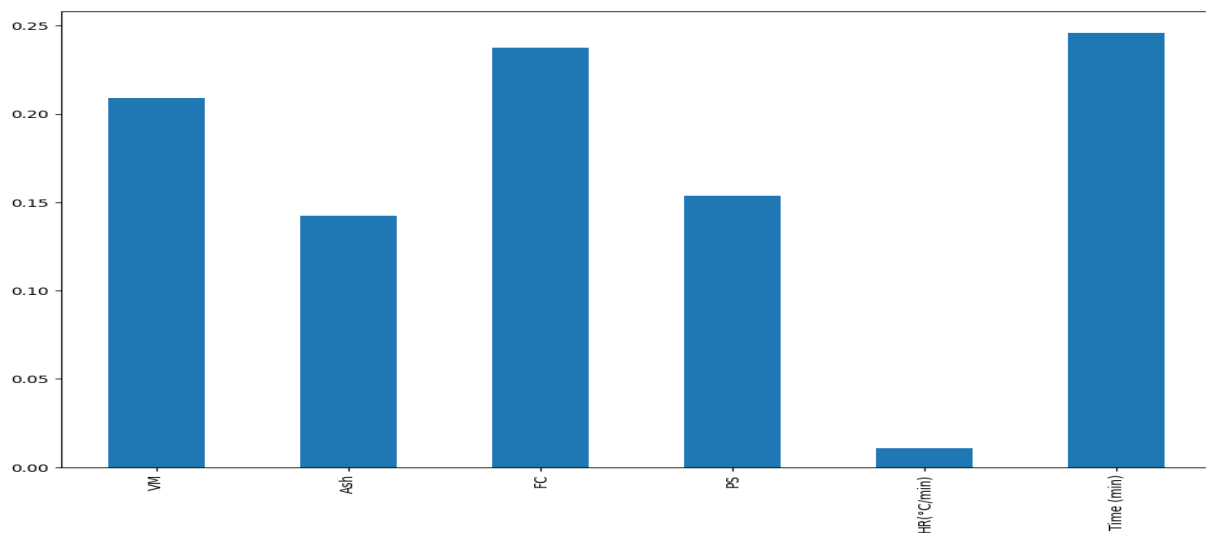
Figure 4.22 Feature Importance plot for HHV prediction

4.4 Experimental Validation

Following the establishment of machine learning models tailored to each specific use case, the primary objective is to validate the obtained results and quantify the disparity between the predicted values and those derived from experimental data. This validation process serves to assess the accuracy and reliability of the predictive models by comparing their outputs against real-world observations. The biochar was synthesized using MAP conducted in a customized microwave reactor at power input of 800 W for 20 min, heating rate of 10 °C/min, 20 °C/min and 30 °C/min along with silicon carbide balls were used as microwave susceptor. Inert atmosphere was provided during MAP by using N_2 gas at a flow rate of 0.2 L/min. Table 4.13 shows the result obtained from experiment and from the predictive models.

Table 4.13. Experimental validation data

Feed-stock	VM (%)	Ash (%)	FC (%)	RT (min)	HR (°C/min).	Bio-Char				Bio-Oil			
						Yield (%)		HHV (MJ/Kg)		Yield (%)		HHV (MJ/Kg)	
						Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Water	68.2	6.8	16.4	20	10	41	37.5	25.23	24.1	43.23	34.5	27.45	22.5
hyacin	69.1	7.23	16.7	20	20	39.94	49.1	26.12	25.3	45.14	37.2	28.78	34.1
th	67.4	6.21	16.1	20	30	37.56	44.5	27.1	27.9	46.96	39.1	29.23	24.3

From the above table we can see that the difference between actual yield and predicted yield for bio-char is less with a RMSE of 8.93 and for bio-oil, RMSE is 16.1. Similarly the difference between the actual HHV and predicted HHV is less with RMSE of 0.75 and 6.11 for biochar and biooil respectively.

<div align="right">

Chapter 5

CONCLUSIONS AND FUTURE SCOPE

</div>

5.1 Conclusion

The study developed models to predict biochar/biochar yield and properties of various biomass through microwave pyrolysis. Developing predictive models to quantify biochar and biooil production based on process parameters is a practical and impactful pursuit in sustainable development. A diverse range of accuracy and error metrics are utilized to evaluate the performance of these machine learning algorithms. The findings unveil crucial insights. Various models were used to predict with the aim of minimum input parameters having high R^2 and low RMSE Values.

- When predicting biochar's yield (%), the Gradient Boosting Regression approach achieved only moderate accuracy when utilizing proximate analysis and process parameters.

- For biochar's SSA prediction, the Support Vector Regressor, incorporating ultimate analysis and process parameters, demonstrated the highest accuracy.

- In the case of biochar's HHV, the Gradient Boosting model, combined with ultimate parameters and process parameters, achieved the least prediction error.

- For biochar's predictions, Ash (%) is most important among proximate parameters and RT (min) is important in process parameters.

- Using Adaptive boosting and proximate and process parameters will give the best result for biooil's yield (%).

- For biooil's Viscosity, Decision tree model is emerged as the optimal choice.

- For biooil's predictions, both FC (%) and RT (min) are turned out to be the important feature.

- The random forest model is best in case of biooil's HHV using proximate analysis parameters.

- Tree based model are turned out to be best predictor.

- Using these models, significant time and effort can be saved in the pursuit of identifying optimal process parameters to enhance yield.

- Proximate parameters and process parameters are the best and minimum set for accurate predictions as compared to the ultimate analysis and process parameters.

These outcomes carry significant implications for biochar production. By implementing these predictive models, substantial time and effort can be saved in identifying optimal process parameters to enhance yield. This approach aligns with the broader objectives of sustainable development, contributing to more efficient and environmentally friendly biochar production processes. The study underscores the value of advanced machine learning techniques in addressing crucial challenges in sustainable development and resource optimization.

5.2 Future Scope

- Further exploration is warranted to investigate the impact of various process parameters, such as feedstock to MW susceptor ratio and MW power to feedstock weight, on both the yield and quality of biochar beyond just reaction time and heating rate.
- Exploring sustainable and cost-effective production methods is imperative to mitigate the environmental impact of biochar production and bolster its economic viability.
- Leveraging artificial intelligence and machine learning techniques holds promise for optimizing both the quality and yield of biochar.
- Detailed investigations into the diverse applications of biochar, spanning agriculture, wastewater treatment, and energy production, are essential to ascertain optimal usage conditions.
- Expanding biomass characterization to encompass additional analytical techniques, like spectroscopy, would yield more comprehensive insights into the biomass's chemical structure, serving as valuable prediction parameters.
- Utilizing advanced techniques like pyrolysis-gas chromatography/mass spectrometry can offer deeper understanding of the molecular and isotopic composition of biochar, enhancing the accuracy of its relationship with yield.

REFERENCES

- Angın, D., & Şensöz, S. (2014). Effect of Pyrolysis Temperature on Chemical and Surface Properties of Biochar of Rapeseed (Brassica napusL.). *International Journal of Phytoremediation*, *16*(7–8), 684–693.

- Chowdhury, Z. Z., Karim, M. Z., Ashraf, M. A., and Khalid, K. (2016). "Influence of carbonization temperature on physicochemical properties of biochar derived from slow pyrolysis of durian wood (Durio zibethinus) sawdust," *BioRes.* 11(2), 3356-3372.

- D. Tilman, R. Socolow, J.A. Foley, J. Hill, E. Larson, L. Lynd, et al.(2009) "Beneficial biofuels–the food, energy, and environment trilemma"

- Hemal Chowdhury, Tamal Chowdhury, PrantaBarua(2021), "Biofuel production from food waste biomass and application of machine learning for process management"

- J.C. Solarte-Toro, J.A. González-Aguirre, J.A. Poveda Giraldo, C.A. Cardona Alzate(2021) "Thermochemical processing of woody biomass: A review focused on energy-driven applications and catalytic upgrading"

- Jin Zhang, Junwei Jin, Minyan Wang, Ravi Naidu, Yanju Liu, Yu Bon Man, Xinqiang Liang, Ming Hung Wong, Peter Christie, Yan Zhang, Chengfang Song, Shengdao Shan, Co-pyrolysis of sewage sludge and rice husk/ bamboo sawdust for biochar with high aromaticity and low metal mobility, Environmental Research, Volume 191, 2020, 110034, ISSN 0013-9351

- Kambiz Tahvildari, Yasaman NaghaviAnaraki, Reza Fazaeli, Sogol Mirpanji, Elham Delrish(2015), "The study of CaO and MgO heterogenic nano-catalyst coupling on transesterification reaction efficacy in the production of biodiesel from recycled cooking oil"

- Kumar, R.; Tiwari, P.; Garg, S.(2013), "Alkali transesterification of linseed oil for biodiesel production"

- Kyung-Won Jung, Tae-Un Jeong, Ho-Jeong Kang, Kyu-Hong Ahn, Characteristics of biochar derived from marine macroalgae and fabrication of granular biochar by entrapment in calcium-alginate beads for phosphate removal from aqueous solution, Bioresource Technology, Volume 211, 2016, Pages 108-116, ISSN 0960-8524

- Lei Luo, Chuang Xu, Zien Chen, Shuzhen Zhang, Properties of biomass-derived biochars: Combined effects of operating conditions and biomass types, Bioresource Technology, Volume 192, 2015, Pages 83-89, ISSN 0960-8524

- Lertsathapornsuk, Pairintra, Aryusuk, Krisnangkura(2008), "Microwave assisted in continuous biodiesel production fromwaste frying palm oil and its performance in a 100 kW diesel generator"

- Leung, D, Guo, Y(2006), "Transesterification of neat and used frying oil: Optimization for biodiesel production"

- Ling Zhao, Wei Zheng, Xinde Cao, Distribution and evolution of organic matter phases during biochar formation and their importance in carbon loss and pore structure, Chemical Engineering Journal, Volume 250, 2014, Pages 240-247, ISSN 1385-8947

- M. Sekar, T. Mathimani, A. Alagumalai, N.T.L. Chi, P.A. Duc, S.K. Bhatia, et al.(2021)"A review on the pyrolysis of algal biomass for biochar and bio-oil - Bottlenecks and scope"

- M.E. Himmel, S.Y. Ding, D.K. Johnson, W.S. Adney, M.R. Nimlos, J.W. Brady, et al.(2007) "Biomass recalcitrance: engineering plants and enzymes for biofuels production"

- Min-Yuan Cheng, Doddy Prayogo(2012), "Optimizing Biodiesel Production from Rice Bran Using Artificial Intelligence Approaches"

- Monzur Alam Imteaz, A.B.M. Sharif Hossain, Maryam Bayatvarkeshi (2022), "A mathematical modelling framework for quantifying production of biofuel from waste banana"

- Patrick Boakye, Hai Nguyen Tran, Dae Sung Lee, Seung Han Woo, Effect of water washing pretreatment on property and adsorption capacity of macroalgae-derived biochar, Journal of Environmental Management, Volume 233, 2019, Pages 165-174, ISSN 0301-4797

- Q. Tang, Y. Chen, H. Yang, M. Liu, H. Xiao, Z. Wu, et al. (2020)"Prediction of Bio-oil Yield and Hydrogen Contents Based on Machine Learning Method: Effect of Biomass Compositions and Pyrolysis Conditions"

- Qiyong Xu, Siqi Tang, Jingchen Wang, Jae Hac Ko, Pyrolysis kinetics of sewage sludge and its biochar characteristics, Process Safety and Environmental Protection, Volume 115, 2018, Pages 49-56, ISSN 0957-5820

- Rouf Ahmad Dar, Mudasir Yaqoob, Manisha Parmar, Urmila Gupta Phutela(2019), "Biofuels from Food Processing Wastes"

- Saba Yavari, Amirhossein Malakahmad, Nasiman B. Sapari, Sara Yavari, Sorption properties optimization of agricultural wastes-derived biochars using response surface methodology, Process Safety and Environmental Protection, Volume 109, 2017, Pages 509-519, ISSN 0957-5820

- Saletnik, B., Zagula, G., Bajcar, M., Czernicka, M., & Puchalski, C. (2018). Biochar and Biomass Ash as a Soil Ameliorant: The Effect on Selected Soil Properties and Yield of Giant Miscanthus (Miscanthus x giganteus). *Energies*, *11*(10), 2535.

- Sandler, L. N., & Gibson, K. A. (2019). A call for weed research in industrial hemp (Cannabis sativa L). *Weed Research*, *59*(4), 255–259.

- Schluttenhofer, C., & Yuan, L. (2017). Challenges towards Revitalizing Hemp: A Multifaceted Crop. *Trends in Plant Science*, *22*(11), 917–929.

- Seo, D.K.; Kim, Y.H.; Eo, Y.D.; Park, W.Y.; Park, H.C.(2017), "Generation of Radiometric, Phenological Normalized Image Based on Random Forest Regression for Change Detection".

- Yang, H., Yan, R., Chen, H., Lee, D. H., & Zheng, C. (2007). Characteristics of hemicellulose, cellulose and lignin pyrolysis. *Fuel*, *86*(12–13), 1781–1788. https://doi.org/10.1016/j.fuel.2006.12.013

- Yousaf, B., Liu, G., Wang, R., Zia-ur-Rehman, M., Rizwan, M. S., Imtiaz, M., Murtaza, G., & Shakoor, A. (2016). Investigating the potential influence of biochar and traditional organic amendments on the bioavailability and transfer of Cd in the soil–plant system. *Environmental Earth Sciences*, *75*(5).