ISSN: 0974-5823 Vol. 6 No. 3 October-December, 2021

International Journal of Mechanical Engineering

Fault Classification in Boiler Drum Using SVM and KNN Prediction Algorithms

Swetha R Kumar*
Department of Instrumentation and Control Systems Engineering, PSG College of Technology, Coimbatore, India

Megalai E
Department of Instrumentation and Control Systems Engineering, PSG College of Technology, Coimbatore, India

Ponkamali P
Department of Instrumentation and Control Systems Engineering, PSG College of Technology, Coimbatore, India

Phavithraa Devi B
Department of Instrumentation and Control Systems Engineering, PSG College of Technology, Coimbatore, India

Gayathri R
Department of Instrumentation and Control Systems Engineering, PSG College of Technology, Coimbatore, India

Kamalakavitha J
Department of Instrumentation and Control Systems Engineering, PSG College of Technology, Coimbatore, India

D Jayaprasanth
Department of Instrumentation and Control Systems Engineering, PSG College of Technology, Coimbatore, India.

Abstract

Boilers are considered as the most important part for the majority of industries particularly in thermal power plant. It is important to improve the performance for a safe and efficient operation. Sudden fault in the boiler may affect the boiler turbine generators chain and it might cause tripping of plant. A fault in a boiler unit causes an enormous loss in the operating revenue. Hence to enhance availability and reduce the shutdown, the application of fault prediction will play a great role in early identification of faults. The performance of the boiler has to be monitored continuously for proper function of plant and to minimize the risk. Using artificial intelligence, the faults are identified much earlier with the help of the input parameters. The preprocessed data are used as input data for the system model which predicts the faults. This paper aims to predict the faults in boilers in power plant using supervised algorithms like support vector machine and K – Nearest Neighbor algorithm. Further the analysis of the aforementioned algorithms and their performance on the given dataset is studied.

Keywords—Support Vector Machine, K Nearest Neighbor, Principal Component Analysis, Boiler drum.

I. INTRODUCTION

Thermal power station is the most traditional wellspring of electric power. Essentially, a thermal power plant comprises of different subsystems like generator, boiler, turbine and so forth. There are various faults which influence the proficiency of boiler. Boiler drum is one of the quintessential segments of a thermal power plant, any deficiency or mistake will cause a decrease in effectiveness. Henceforth there is an extraordinary need to appropriately distinguish and group the fault ahead of time, to stay away from any decrease in effectiveness. The productivity of the power plant depends a lot on boiler drum proficiency. Henceforth, it is important to complete the upkeep of the boiler at customary stretches to stay away from any undesirable events. Also, the evaporator prompts several issues like agglomeration, slagging, fouling, acidic embrittlement, weariness disappointment, and hot erosion. In addition, the fireside of boilers endures the extreme damage due to faults.

Several machine learning algorithms for fault detection and classification is studied by researchers across. Rony et al [1] developed a improved fault detection algorithm in two ways: first, it presents a validated boiler emulator model that can be used to generate simulated data for rare fault

conditions, and second, it explores machine learning models for fault detection using points typically monitored by Building Automation Systems (BAS). A selection of common classification algorithms was tested for their ability to distinguish between normal operation and each fault, Naïve Bayes (NB), Decision Trees (DT), and Random Forest (RF) using 500 trees. The algorithms were programmed in R which was used for model training, testing, evaluation, and result visualization.

Elisa Guelpa and Vittorio Verda in their work proposed a methodology which detects the level of fouling in heat exchanger. The mass flow rate on the primary side and the temperatures on both sides of the heat exchanger are taken as input, Evaluation is difficult due to the rawness of the data gathered and the variable operating conditions, which are adjusted on the basis of the external temperatures and set-points. [2].

Jungwon et al [3] proposed a fault isolation method via Classification and Regression Tree (CART). In this work, binary classification trees are constructed by applying the CART algorithm to a training dataset which is composed of normal and faulty samples for classifier learning. Then, to perform faulty variable isolation, variable importance values for each input variable are extracted from the constructed trees. Furthermore, this method is based on the nonparametric CART classifier, can be applicable to nonlinear processes.

It is observed that, the dynamic model which is developed using NARX neural network architecture has the better prediction of boiler drum level when compared with the static model. There by comparing the results the error is detected if any [4]

II. FAULTS IN BOILER DRUM

One of the principle issues in the boilers are scaling in evaporator is brought about by degradations which are being hastened out of the water straightforwardly on warmth move surfaces. Scaling in boiler is caused by impurities which are being precipitated out of the water directly on heat transfer surfaces. Evaporation in a boiler causes impurities to concentrate which interferes with heat transfers and may cause hot spots, leading to local overheating.

Scaling mechanism is the exceeding of the solubility limits of mineral substances due to elevated temperature and solids concentration at the tube/water interface. The deposition of crystalline precipitates on the walls of the boiler interferes with heat transfer and may cause hot spots, leading to local overheating.

Fig.1: Scaling

The second principle issue is fouling. When the deposit is built up by condensed materials, forming a dry deposit, generally in the convective section, then it is called fouling. Fouling is most likely on the preheater and super heater surfaces where the flue gas temperatures are much lower than in the combustion zone. Deposition in a boiler depends mainly on the amount and type of inorganic materials present in the coal being used.

Fig. 2: Fouling

Another fault in boiler drum is excess air. With boiler combustion, if some excess air is not added to the combustion process, unburned fuel, soot, smoke, and carbon monoxide exhaust will create additional emissions and surface fouling. From a safety standpoint, properly controlling excess air reduces flame instability and other boiler hazards. Even though excess air is needed from a practical standpoint, too much excess air can lower boiler efficiency. So, a balance must be found between providing the optimal amount of excess air to achieve ideal combustion and prevent combustion problems associated with too little excess air, while not providing too much excess air to reduce boiler efficiency.

Fault detection and classification is needed for safe operation and ideal control of mechanical cycles other than decrease of upkeep and activity costs.

III. ALGORITHM WORKFLOW

Artificial Intelligence (AI) algorithms are programs (math and rationale) that change themselves to perform better as they are presented to more information. Machine learning is one of the most important applications of Artificial Intelligence (AI). It is widely used nowadays because it provides the ability to the system to automatically learn from the experience instead of coding the algorithm explicitly. Machine learning workflow involves five major

steps which are generally followed in every application as shown in fig 4.

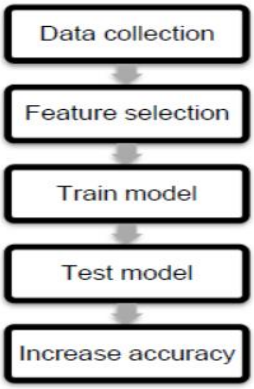

Fig 3: Workflow in Machine learning

Data collection is the first and foremost step in machine learning. Collection of data can be either collected from an existing database or it can be collected in real time using an IOT system.

Feature selection is a core concept in machine learning workflow which hugely impacts the model's performance. Generally, feature selection is a process in which the numbers of input variables are reduced while developing a predictive model. If the features selected are irrelevant to the output variable or if it is partially irrelevant the accuracy of the system will be highly reduced and also prediction will be of increased error.

Major benefits of performing feature selection are,

1. Reduction in over fitting, because possibilities of training the model with noise is highly reduced.
2. Increase in Accuracy, since there is a reduction in misleading data.
3. Reduction in Training time, since there is a reduction in amount of data and complexity of algorithm.

Three major feature selection techniques implemented in this paper are,

1. Univariate selection: This is a statistical method that helps to find the feature with strongest relationship with the output variable. This selection method examines each and every feature individually in order to determine its relation.
2. Feature importance: This technique helps us to find the individual importance of each and every feature. This feature is inbuilt in Tree based classification algorithms but if other

algorithms are used it must be additionally inserted.

3. Correlation matrix with Heatmap: Correlation generally indicates the relationship between the feature and output variable, whereas heatmap helps us to visualize the relationship.

Thus feature selection is an important step which reduces over fitting, and also increases accuracy. Along with this reduction in dimension is also possible. To train the model SVM and KNN algorithms are used

A. SUPPORT VECTOR MACHINE ALGORITHM

SVM is most commonly used algorithm for classification-based problems because of its less computational power and significant accuracy. Support vector machine which is abbreviated as SVM can be used for both regression and classification-based problems. The main aim of SVM algorithm is to find a hyperplane which will distinctly differentiate the dataset. Hyperplane must be chosen optimally based on the parameter known as Margin. Hyperplane with maximum margin is always selected because only then there will be maximum distance between the data points of different classes.

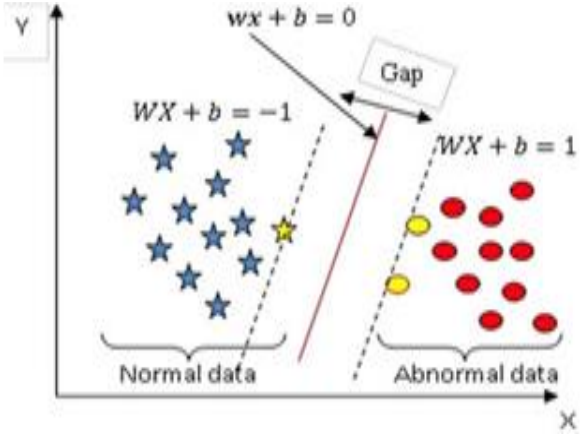

Fig.4 Support vector machine algorithm

Hyperplanes can be of different dimensions based on the number of features considered. Kernel function is a term commonly used in SVM algorithm which helps us to convert non-linear decision surface to linear decision surface with the help of few mathematical manipulation.

Important parameters that affect the accuracy in SVM algorithm are: Regularization, Gamma, Kernel function, Hyperplane. The SVM attempts to set an on the right track limit between two interesting classes, and mastermind it with the goal that the edge is enhanced. All things considered, the SVM endeavors to mastermind the cutoff with the ultimate objective that the distance between the breaking point and the nearest data point in each classes maximal. The hyperplane is then situated in this edge between the two centers (Fig. 4).

To show this, training samples are considered as

$$X \in Rn$$
$$Y \in \{-1,1\} \quad (1)$$

The decision function will be

$$F(X) = \text{sign}(\langle W, X\rangle + b)$$
$$W \in R_n \quad (2)$$
$$b \in R$$

The objective function is characterized in a manner to expand, D, the distance between two planes or similarly minimize vector W

$$\text{MAXIMIZE } D = \frac{2}{||W||} \quad (3)$$

The enhancement issue is addressed by presenting imperatives (5) into the target work by means of lagrangian multipliers.

$$Y_i\left(\langle \vec{W}, \vec{X}\rangle + b\right) \geq 1 \qquad (4)$$

The lagrangian is

$$= \frac{1}{2}\sum_{i-1}^{n} w_i^2 - \sum_{i=1}^{N^{A_{p(\vec{w},b,\vec{a})}}} a_i \left(Y_i\left(\overrightarrow{W}.\overrightarrow{X_i} + b\right) - 1\right) \qquad (5)$$

Where n shows the component of framework information, N is number of preparing tests and a_i is Lagrangian multiplayer acquired from

$$a^* = \text{argmin}_\alpha \frac{1}{2}\sum_{i=1}^{N}.\sum_{j=1}^{N} \alpha_i\alpha_j Y_i Y_j \langle X_i, X_j\rangle - \sum_{k=1}^{N}\alpha_k \qquad (6)$$

Finally, the optimal values of W* and b* are obtained by Lagrangian multiplier which are

$$W^* = \sum_{i=1}^{N} \alpha_i Y_i \vec{X}_i \qquad (7)$$

$$b^* = \frac{1}{2}\langle W^*, X_r + X_S\rangle \qquad (8)$$

where, X_r and X_s are data of two classes. Eventually the separating function will have the form of

$$F(\overrightarrow{X}) = \text{sign}(\langle \overrightarrow{W^*}, \vec{X}\rangle + b^*) \qquad (9)$$

which is known as hard characterization. Information isn't in every case straight detachable. Considering this error rate, the target capacity to be limited will be changed to

$$\frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}\varepsilon_i \qquad (10)$$

which is called soft-margin SVM. Therefore having (10), the separable function will have

$$Y_i\left(\overrightarrow{WX} + b\right) \geq 1 - \varepsilon_i \qquad (11)$$

At the point when C is main parameter, delicate edge SVM is comparable to hard one and with little C, we concede misclassification in the preparation information arranged by having w-vector with little standard [10]. In the new space, information can be straight ordered and isolating condition will have the structure

$$F(\overrightarrow{X}) = \text{sign}\left(\sum_{i=1}^{N}\alpha_i Y_i K(\overrightarrow{X_\iota}, \vec{X})\right) \qquad (12)$$

where K(.,.) is a kernel function that should satisfy Mercer's condition [11]. In this work Gaussian kernel function is used

B. K Nearest Neighbour

KNN algorithm is broadly applied in example acknowledgment and information digging for arrangement, which is acclaimed for its straightforwardness and low blunder rate. The guideline of the algorithm is that, if larger part of the k most comparative examples to a question point qi in the component space have a place with a specific classification, at that point a decision can be made that the inquiry point qi fall in this class. Comparability can be estimated by the distance in the component space, so this algorithm is called K Nearest Neighbour algorithm. A train informational collection with exact characterization names ought to be known toward the start of the algorithm. At that point for a question information qi, whose mark isn't known and which is introduced by a vector in the component space, ascertain the distances among it and each point in the train informational index. In the wake of arranging the consequences of distances computation, choice of the class mark of the test point qi can be made by the name of the k closest focuses in the train informational collection.

Each point in d-dimensional space can be expressed as a d-vector of coordinates, such as:

$$p = (p_1 p_2, \ldots\ldots, p_n) \qquad (14)$$

The distance between two focuses in the multidimensional element space can be characterized from multiple points of view. Utilizing Euclidean distance is normally to be the most standard technique, that is:

$$\text{dist}(p, q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \qquad (15)$$

Alternatively, Manhattan distance can also be used as:

$$\text{dist}'(p, q) = \sum_{i=1}^{n}|p_i - q_i| \qquad (16)$$

The nature of the train informational index straightforwardly influences the order results. Simultaneously, the decision of boundary K is likewise vital, for various K

IV. Selection Of Parameters For Svm And Knn Algorithms

The dataset is taken from IEEE Data Port [1], which consists of 27281 data. In the class column, the fault is categorized as three types namely scaling, fouling and excess air which are described above.

The inputs are

- Fuel_Mdot – Fuel flow rate [kg/s]
- Tair – Ambient air condition [K]
- Treturn – Temperature of water entering the boiler [K]
- Tsupply – Temperature of water in the boiler [K]
- Water_Mdot – Boiler loop flow rate [kg/s]

The outputs are following Fault category

- 0 - Lean
- 1 - Nominal
- 2 - Excess air
- 3 - Fouling

- 4 - Scaling

The parameters make a major role that has been taken place from that to extract the maximum accuracy from a classifier it is essential to choose appropriate values for the parameters that plays a major role in the design of decision boundary of the model. In the case of SVM classifier one such important parameter is the regularization parameter 'C'. For the purpose of choosing the optimum regularization parameter for the SVM classifier, a graph is plotted between a range of C values and the accuracy of the model corresponding to it as shown in Fig 5

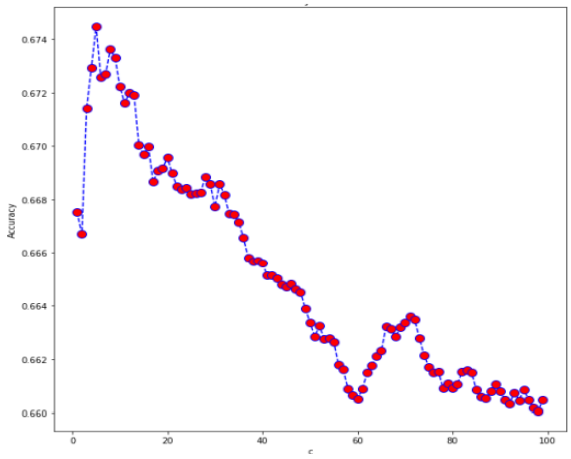

Fig.5 'C' value vs Accuracy of SVM classifier

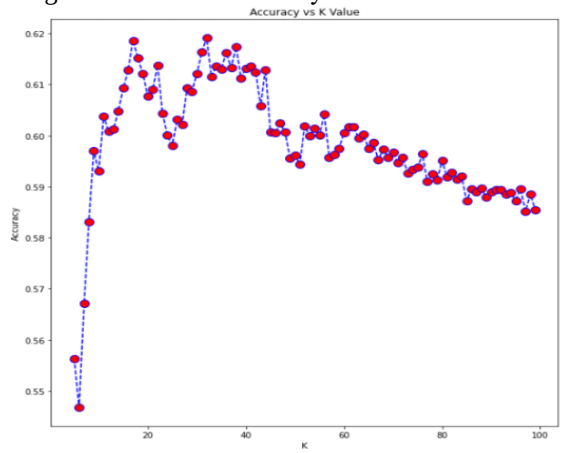

Fig.6 'K' value vs Accuracy of KNN classifier

From Fig 5, it is clear that maximum accuracy is achieved for C = 5 and so the value of 'C' is fixed to be 5. In order to choose the optimum nearest neighbor value for the KNN classifier, a graph is plotted between a range of K values and the accuracy of the model corresponding to it as shown in Fig 6.

V. Result And Discussion

In this work the dataset is taken from IEEE Data Port, which consists of 27281 data. In the class column, the fault is categorized as three types namely scaling, fouling and excess air. The parameters and responses are optimized from the Support vector machine algorithm and KNN algorithm to determine which parameter influences more in the efficiency and defects in the boiler. By the SVM and KNN algorithms the efficiency of the algorithms is calculated by comparing their feature selection is observed. The decision boundary region as well as the accuracy of SVM and KNN algorithms is compared.

For the three responses scaling, fouling and excess air the best featuring parameter is checked by using two methods to find optimization of the data set to determine which parameter affects the efficiency of the boiler. From the output from univariant score it is clear that the highest scoring as well as statistically high priority feature in the dataset is TSupply (Temperature of the water in the boiler). Which affect the efficiency and fault of the boiler shows in the Fig 7.

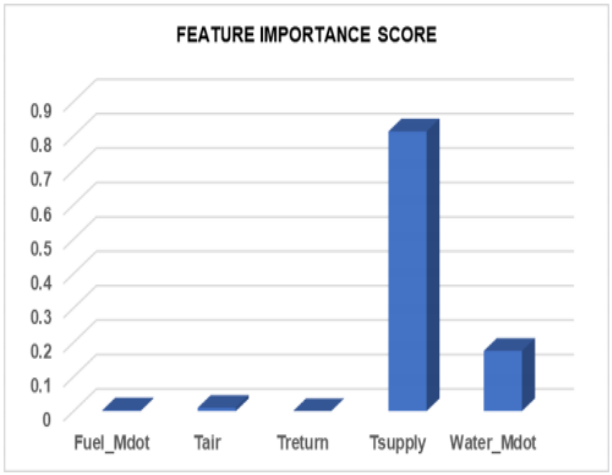

Fig.7 Feature importance score

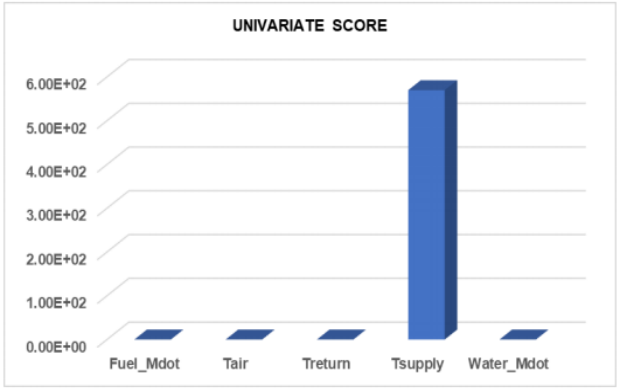

Fig.8 Univariate score

To verify the deciding parameter of the boiler. the Tsupply is the deciding factor of the efficiency of the boiler from univariate score. So, the feature impotence score analysis has been taken for the analysis again the Tsupply has been taken the leading score point which shows from the graph. Thus, the leading score graph from the feature importance score shows the tsupply is deciding factor from both the analysis.

Being a classification algorithm, the accuracy of the method should be presented as routine. Figure gives us the accuracy of the algorithm. The accuracy of the two methods shows little difference, for the principle of them is just the

same. In KNN algorithms, the final result depends entirely on the quality of the data set. The slight difference of the results is due to sorting step for the same distances between the query point and that from train set with different category label was cut by the parameter K differently because of unstable sorting algorithms.

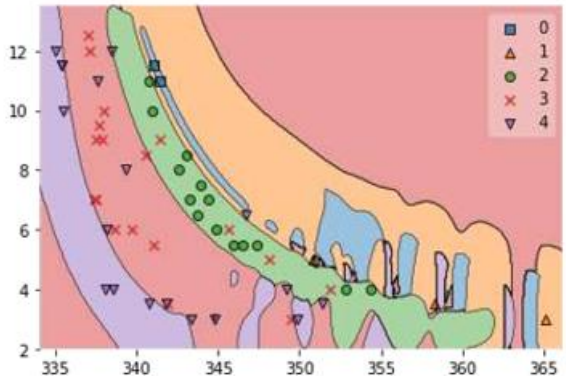

Fig.9 SVM decision region boundary

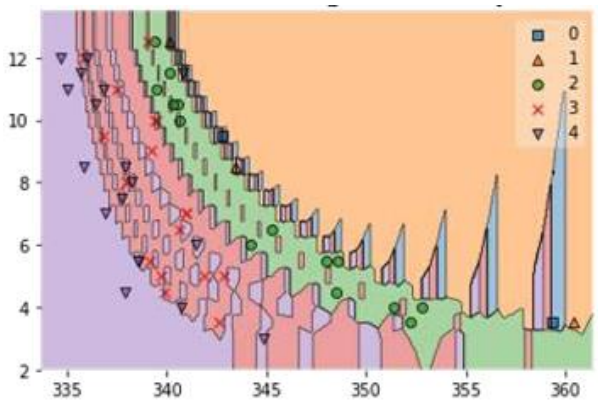

Fig.10 KNN decision region boundary

The above image shows the boundary regions of the SVM classifier. It can be seen that the many of the test data are placed in wrong boundary regions resulting in low accuracy of the classifier Fig.9 shows the boundary regions of the KNN classifier. It can be seen that most of the test data are placed in correct boundary regions and hence there is a higher accuracy when compared to the SVM classifier.

The sample output from the SVM and KNN classifier is shown in the table below

Table.1 SVM and KNN Prediction vs actual

S.no	Tsupply	Water_ Mdot	SVM		KNN	
			Actual class	Predicted class	Actual class	Predicted class
1	355.1203	4	1	1	1	1
2	344.9716	3.5	4	3	4	4
3	348.2677	5	2	2	2	2
4	358.2767	3	3	2	3	2
5	353.5806	4.5	3	0	3	3

Before feature selection both the classifiers have similar accuracy range whereas after feature selection there is a clear distinction in their accuracies. Fig 11 shows that the accuracy of the KNN classifier (94%) is higher than the accuracy of the SVM classifier (74%).

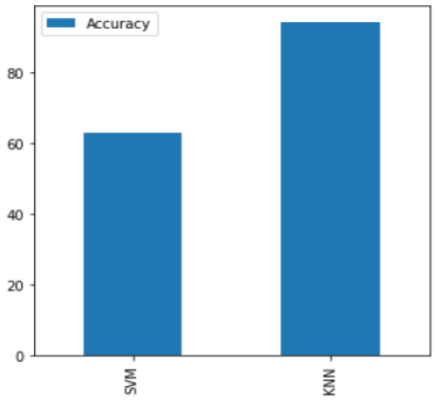

Fig.11 Comparison of SVM and KNN classifier

VI. CONCLUSION

The identification and classification of faults that can occur in a boiler is quite essential for efficient generation of power. By employing SVM and KNN algorithms the faults viz., scaling, fouling and excess air in a boiler drum were classified. By detecting the faults, the chances of reduction in efficiency is prevented and therefore ensure proper generation of power in the thermal power plant.

- The deciding parameter of the boiler will be the Tsupply which is the response where the boiler depended for its efficiency and its defects

- The proposed model was able to achieve accuracy of about 94% which is considerably higher. Therefore, the model can serve to predict and classify the faults in a boiler system

REFERENCES

[1] R. Shohet, M. Kandil and J. McArthur, "Machine learning algorithms for classification of boiler faults using a simulated dataset", IOP Conference Series: Materials Science and Engineering, vol. 609, p. 062007, 2019.

[2] E. Guelpa and V. Verda, "Automatic fouling detection in district heating substations: Methodology and tests", Applied Energy, vol. 258, pp. 114059, 2020.

[3] Yu, J. Jang, J. Yoo, J. Park and S. Kim, "A Fault Isolation Method via Classification and Regression Tree-Based Variable Ranking for Drum-Type Steam Boiler in Thermal Power Plant", Energies, vol. 11, no.5, p. 1142, 2018.

[4] B. S. T. Selvi, D. Kalpana and T. Thyagarajan, "Modeling and prediction of boiler drum in a thermal power plant ", 2017 Trends in Industrial Measurement and Automation (TIMA), Chennai, pp. 1-6, 2017.

[5] A. Singh, V. Sharma, S. Mittal, G. Pandey, D. Mudgal and P. Gupta, "An overview of problems and solutions for components subjected to fireside of boilers", International Journal of Industrial Chemistry, vol. 9, no. 1, pp. 1-15, 2017.

[6] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa and A. Mueller, "Scikitlearn: Machine Learning in Python", Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.

[7] "Water Treatment Solutions," Lenntech Water treatment & purification. [Online]. Available:.

[8] J. Mahato, "What is boiler: Types of boiler: How does a steam boiler work," Coal handling plants, 11 Apr 2020. [Online]. Available: https://www.coalhandlingplants.com/boiler-in-thermal-power-plant/

[9] "Combustion Efficiency and Excess Air," Engineering ToolBox. [Online]. Available: https://www.engineeringtoolbox.com/boiler-combustion-efficiency-d_271.html

[10] Ravi, "Components and Working of Thermal Power Plants," Electrical Article, 02-Mar2019. [Online]. Available: http://electricalarticle.com/components-working-thermalpower-plants/