INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

Machine Learning Approach for Detecting Anomalies in Industrial Energy Consumption

1st Priyadharshini K
Department of Advanced Computing & Analytics
Vels Institute of Science,Technology & Advanced Studies
Chennai, India
priyanidhu15@gmail.com

2nd Dr.T. Sree Kala, Professor
Department of Advanced Computing & Analytics
Vels Institute of Science,Technology & Advanced Studies
Chennai, India
sreekalatm@gmail.com

***Abstract*-Industrial energy usage is at the heart of increasing efficiency, reducing costs, and promoting proper usage of sustainable energy. When there is a deviation in energy usage, it may indicate issues such as machine failures, leaks, or inefficient usage. Detecting these issues early can help reduce machine downtime and maintenance costs. This project aims to create an anomaly detection system using machine learning algorithms based on a simulated dataset, since acquiring live data from industries is difficult. We use algorithms like Isolation Forest, One-Class SVM, and Gradient Boosting to detect anomalies in energy usage. We analyze time-series features like power usage, voltage, and other signals to detect anomalies. A dashboard is also provided for easy visualization of energy usage and anomalies. The results show that this method is helpful for predictive maintenance and increasing efficiency in industries. To sum up, machine learning can be very helpful in anomaly detection in this domain.**

I. INTRODUCTION

It is critical for industries to monitor their energy use when dealing with high-capacity production areas like thermal power plants. A thermal plant includes multiple types of machinery/processing equipment; many of these machines generate large amounts of when dealing with high-capacity production areas like thermal power plants. A thermal plant includes multiple types of machinery/processing equipment; many of these machines generate large amounts of data. Each machine will have at least a few key indicators (e.g.: temperature, pressure, speed, and output energy) that must be monitored continuously to ensure that the machinery operates smoothly, that production costs are minimised, and to avoid having a piece of machinery fail. data. Each machine will have at least a few key indicators (e.g.: temperature, pressure, speed, and output energy) that must be monitored continuously to ensure that the machinery operates smoothly, that production costs are minimised, and to avoid having a piece of machinery fail.

It is critical for industries to monitor their energy use when dealing with high-capacity production areas like thermal power plants. A thermal plant includes multiple types of machinery/processing equipment; many of these machines generate large amounts of data. Each machine will have at least a few key indicators (e.g.: temperature, pressure, speed, and output energy) that must be monitored continuously to ensure that the machinery operates smoothly, that production costs are minimised, and to avoid having a piece of machinery fail. Industrial control systems (ICS) play an important role in managing the electricity grid, as well as water treatment plants and other critical infrastructure. ICS typically use programmable logic controllers (PLC) and SCADA as part of their control process, in addition to other tools, to help maintain stable operations.

This paper discusses the application of machine learning techniques to the development of anomaly detection systems for energy consumption in the industrial sector. The recent increase in datasets/availability and the availability of higher computational processing power has allowed machine learning techniques to provide an effective and efficient method of detecting anomalies within industrial energy consumption data. Also, these techniques can help improve the efficiency of machinery and, at the same time, reduce the number of system failures when used within a large-capacity production area.

Using the Isolation Forest algorithm, we are able to detect anomalies within the data collected within the case study, allowing for the identification of potential operational issues that could lead to machinery operating at less than optimal performance or potentially causing a fault in a piece of machinery. This paper demonstrates using the Isolation Forest algorithm for anomaly detection provides a basis for real-time monitoring of energy usage, improving energy efficiency, and providing predictive maintenance.

II. LITERATRUE SURVEY

The increasing number of industrial automation systems and smart energy systems are requiring more efficient methods to monitor energy usage. Consequently, machine learning and artificial intelligence (AI) methods are now being employed by researchers to detect patterns of abnormal energy use and locate malfunctions or faults in

power generation systems. The Isolation Forest algorithm is one of the best algorithms for detecting anomalies; it works by isolating anomalous data points based on their unique features [1].

Different types of machine learning algorithms, including support vector machines (SVM), k-nearest neighbour (KNN), logistic regression (LR), and multilayer perceptron (MLP), have been used to identify unusual patterns of energy consumption. The random forest method is one of the most accurate machine learning models, achieving approximately 90 – 95% accuracy in intelligent power systems. In addition, AI-based techniques can successfully process and analyze large volumes of data.

More recently, autoencoder techniques have achieved very high detection rates (90 – 97%) of anomalous behaviours in power systems. Advanced techniques such as gradient boosting and deep learning have achieved a high level of performance in identifying anomalies from industrial data [4][5]. However, issues still exist with machine learning techniques for anomaly detection; specifically, data availability, large number of variables/dimensions, and explainability of machine learning models. Thus, it is vital that reliable systems for detecting anomalies are developed using machine learning [2][3].

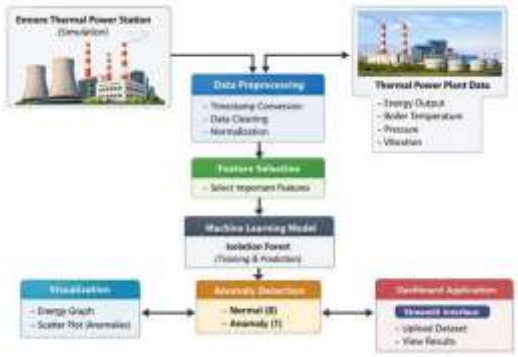
III. PROPOSED METHODLOGY

The initial step is to gather information pertaining to how much energy is consumed by the index of industries, similarly as with the operation of a thermal power plant (in this example, Ennore Thermal). Once the information has been gathered, the information will be processed by establishing a methodical catalogue via the three analytical methods outlining how to appropriately clean, align and normalize the data for purposes of correlation/comparison and establish consistency through a series of procedures.

The data set is then passed into the Isolation Forest model, allowing for the model to receive training that will enable it to differentiate between normal operation conditions and anomalous conditions through identification of anomalies which are defined as points of dissimilarity to the expected norm via training of the Isolation Forest model. The results will then be displayed via graphs through use of the Streamlit platform, allowing users the opportunity to examine patterns of irregularities in terms of energy consumption. An end user of the proposed process will have access to a singular interface that will allow users to upload their data sets, output their models, and access the results in real time.

The emergence of Machine learning as a tool/technology for the purpose of discovering anomalous behaviour is now going to increasingly be a practical application due to the wealth of available datasets and the accessibility of robust computational power. The proposed methodology in this work will leverage the Isolation Forest algorithm to determine anomalies as it relates to energy consumption behaviour [1] while also highlighting numerous operational challenges associated with employing the proposed solution in a real world setting, such as a construction site, water utility facility, etc. [1].

IV. ARCHITECTURE DIAGRAM



V. VARIOUS METHODOLOGY

Within this project, an intelligent anomaly detection system is developed, which keeps an eye on the energy consumption within the industrial setup of thermal power plants. This is done through the application of machine learning techniques. Data collected from the Ennore Thermal Power Station is used for the purpose of identifying any anomaly within the energy consumption, which could point towards any possible issues within the system.

A) Data Collection

- Timestamp (time interval of the data recorded)
- Unit Load (MW)
- Boiler Temperature (°C)
- Steam Pressure (bar)
- Coal Flow Rate (tonnes/hour)
- Turbine Speed (RPM)
- Cooling Water Temperature (°C)
- Energy Output (MW)

B) Data Preparation

Preprocessing activities include:

- Converting the timestamp data into a proper data type

- Using the timestamp as the index for the dataset
- Checking for any missing data within the dataset
- Creating a statistical summary of the data

D) Machine Learning-Based Anomaly Detection

A physical process within a control system is maintained within the boundaries defined for that process. This means that every state variable has its own boundary, and the controller will work towards ensuring that there is no drift outside these boundaries. For instance, the water level within a tank has to be maintained within a certain low and a certain high level. Over a period of time, all state variables will change based on the process dynamics. For instance, in the case of ultrafiltration, cleaning has to be done every 30 minutes or so in order for the pressure drop to be within acceptable limits.

Anomaly detection models can be considered a black box that receives data in real-time from a plant, for instance, the SwaT plant. In essence, there are two broad classes of machine learning models that are used for anomaly detection in ICS:
models that analyze the relationships between feature vectors in order to detect anomalies, and models that make predictions about how the ICS should be performing and detect any anomaly based on how the ICS is performing compared to the prediction.

E) Model Evaluation

A physical process within a control system is maintained within the boundaries defined for that process. This means that every state variable has its own boundary, and the controller will work towards ensuring that there is no drift outside these boundaries. For instance, the water level within a tank has to be maintained within a certain low and a certain high level. Over a period of time, all state variables will change based on the process dynamics. For instance, in the case of ultrafiltration, cleaning has to be done every 30 minutes or so in order for the pressure drop to be within acceptable limits. Anomaly detection models can be considered a black box that receives data in real-time from a plant, for instance, the SwaT plant. In essence, there are two broad classes of machine learning models that are used for anomaly detection in ICS: models that analyze the relationships between feature vectors in order to detect anomalies, and models that make predictions about how the ICS should be performing and detect any anomaly based on how the ICS is performing compared to the prediction.

G) Anomaly Reporting

There are a few distinct steps in the process of creating an anomaly detector system. These steps are model development, validation, and testing; then deployment, fine-tuning, and finally operation, with re-tuning as and when necessary. Figure 1 shows a flowchart of these steps. In the development phase, a decision must be made on which approach to use in creating a model, which will finally become an anomaly detector. There are a number of possible approaches that can be used in creating a model that can finally become an anomaly detector.

Finally, the system will produce a list of identified anomalies along with the corresponding operational parameters, which can be used by engineers to understand why the energy consumption is behaving in an unusual manner and take appropriate steps for resolution. The Anomaly Reporting module can help in understanding the possible causes for certain issues, which can be identified as follows:

- Overheating of the boilers
- Abnormal levels of steam pressure
- Turbine speeds behaving in an unusual manner
- Sudden fall in energy production

All these issues can be identified early on with the help of the proposed system, making the operation of a thermal power plant even more efficient.

VI. INPUT

The system receives real-world data from a thermal power plant. The data set contains a range of parameters, all of which are related to energy production. The data is then processed using a machine learning model, which detects abnormal or unusual activity within the plant. The data set contains several variables, including a time stamp for each reading, showing when the reading was taken. The time stamp also indicates the date and time of each reading. The second variable is the amount of energy produced, measured in megawatts. The third variable is the temperature of the boiler, measured in degrees Celsius. The temperature of the boiler indicates how hot it is during operation. The pressure of the boiler is also measured, as is the mechanical vibration of the equipment. A variable is also provided for abnormal readings.

VII. PSEUDO CODE AND IMPLEMENTATION

Begin

1) Import necessary libraries

- Streamlit
- Pandas
- Matplotlib
- StandardScaler

- Isolation Forest

2) Display title/subtitle of application

3) User to upload file (dataset)
IF user uploads a file ,
- Read file into dataframe
- Print out first n rows of file
- Change 'Timestamp' column's values to be in datetime format
- Use 'Timestamp' column as the index

4) Feature selection
- Remove anomaly column
- Store remaining features in input variable X

5) Normalization of data
- Scale input variable using StandardScaler to obtain input variable X_scaled

6) Train model — use Isolation Forest algorithm to develop trained version of model using input variable X_scaled

7) Predict anomalies using the trained model

- Convert predictions to create new column for anomalies

IF prediction = -1 then anomaly = 1
ELSE anomaly = 0
- Store output in the new column of the dataframe — Predicted_Anomalies

8) Compute statistics
- Find total records
- Find total anomalies
- Find percentage of total anomalies versus total records

9) Show stats — total records, total anomalies, percentage of total anomalies

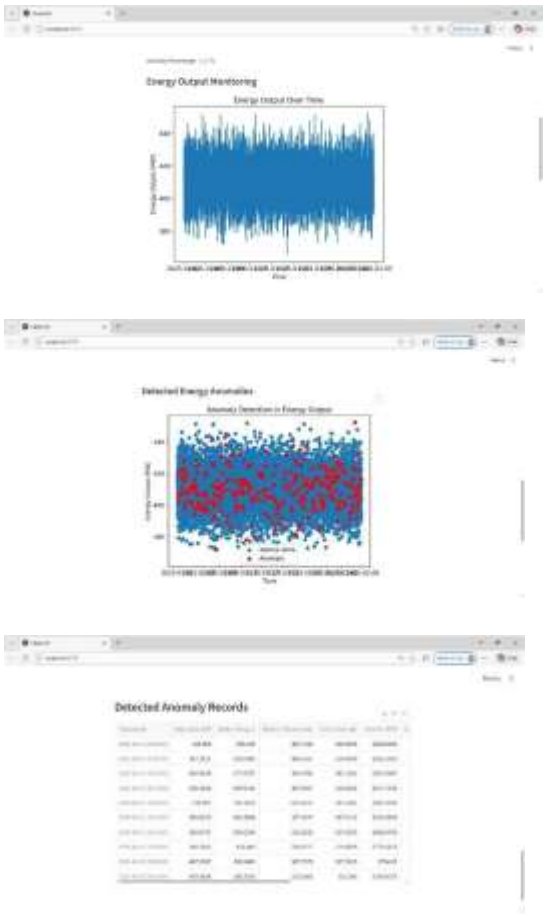
10) Create visualizations
- Plot energy output against time in line graph
- Create scatter graph with:
  - Non-anomalous (normal) data points
  - Highlighting data points determined to be anomalies by coloring them red

11) Display Detected Anomalies
- Action: Filter rows where the value in the Predicted_Anomalies Column is 1

Implementation

This system was created using Python and libraries like Pandas; Scikit-learn; and Matplotlib. First, load and pre-process new data into the correct format so it can be analyzed. Next, identifying key features you would like to capture as input parameters will aid in creating an accurate model when using those specific features for the analysis. Your input parameters should include energy output, boiler temperature/pressure/vibration, and any other data or calculations that are relevant to your analysis. After selecting all of these features, use standardScaler to normalize their values in order to help maintain the accuracy of the models being created with different scales used to represent these features. Finally, apply an isolation forest model to identify anomalies using abnormal patterns found; record the predicted result of each individual sample in your sample dataset into a new column.

VIII. OUTPUT



IX. RESULTS AND DISCUSSIONS

A synthetic dataset was generated to mimic the conditions of the thermal power plant, similar to the Ennore plant. The dataset was generated to include the necessary parameters that are monitored, including the timestamp, energy output, temperature, pressure, and boiler vibration, among others. The data was then ready for analysis by aligning the timestamps and normalizing the data. Anomaly detection was achieved by using the Isolation Forest method, which comes with the Scikit-learn library. The model was able to perform the necessary task of detecting anomalies in the data, differentiating between the normal and abnormal data patterns. The anomalies were also detected, indicating some unusual conditions in the plant, including sudden changes in the energy output.

Finally, to complete the whole process, the Streamlit library was integrated to provide the necessary interface for the entire process, including the anomaly detection, to be integrated into one interface, providing the necessary tool to detect unusual energy consumption patterns, thereby increasing the efficiency of the plant.

X. CONCLUSION

Geared towards analyzing energy usage from the viewpoint of both the output (energy produced) as well as the boiler's performance (temperature, pressure, and vibration), this data has been cleaned and pre-processed to facilitate accurate analyses. In particular, an Isolation Forest Model was developed and implemented using Scikit-learn to detect anomalies within the dataset. The output of the Isolation Forest model has shown to be effective at identifying anomalies associated with industry energy usage. Visualizations of the identified anomalies within the output were created using Matplotlib to assist users in communicating their results in an easy-to-understand fashion. Additionally, a dashboard has been created within the output that allows users to more easily upload their datasets and detect anomalies using a graphical user interface (GUI) with user-friendly graphs and informative, descriptive statistics. As such, this system provides both a straightforward and effective means to monitor energy usage within an industrial environment, as well as identify operational anomalies associated with the operation of a power generation facility. Moreover, these types of systems also allow operators the ability to identify potential failures before they occur, operationally optimize their processes, reduce the amount of wasted energy, and improve their industrial energy management decisions.

XI. REFERENCE

[1] F. T. Liu, K. M. Ting, and Z. H. Zhou, “Isolation Forest,” in *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, 2008, pp. 413–422.

[2] B. Schölkopf et al., “Estimating the Support of a High-Dimensional Distribution,” *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[3] S. Chandola, A. Banerjee, and V. Kumar, “Anomaly Detection: A Survey,” *ACM Computing Surveys*, vol. 41, no. 3, 2009.

[4] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[5] J. H. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine,” *Annals of Statistics*, 2001.

[6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[7] Scikit-learn Documentation. [Online]. Available: https://scikit-learn.org

[8] Streamlit Documentation. [Online]. Available: https://streamlit.io

[9] Pandas Documentation. [Online]. Available: https://pandas.pydata.org

[10] Matplotlib Documentation. [Online]. Available: https://matplotlib.org