Clustering-based Classification Model for failure prediction on power transformers.

Harshil Lakkad
Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology
Pune, India
Harshilpatel98981@gmail.com

Athava Undale
Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology
Pune, India
Atharvaundale60@gmail.com

Akshay Marathe
Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology
Pune, India
Akkim17@gmail.com

Dr. Shivali Wagle
Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology
Pune, India
Shivali.wagle@sitpune.edu.in

Dr. Preksha Parekh
Department of Artificial Intelligence and Machine Learning, Symbiosis Institute of Technology
Pune, India
Preksha.pareek@sitpune.edu.in

Abstract: **Machine maintenance plays a crucial role in the realm of power distribution systems. It helps in preventing downtime of the system and prevents loss of time, money, and energy. Since Distribution Transformers are a vital component of the Power delivery system, prevention of its failure becomes a major task. Here we use Preventive Maintenance using AI models to try to predict if the transformer is going to be burned or not. We have implemented various Machine Learning algorithms for this problem and achieved varying success. The use of multiple models for predicting different factors is a common practice. One of the gaps for predicting failure in real-time is the complexity of multiple factors working simultaneously that can cause the failure of the machine.**

Index Terms: **Machine Maintenance, Preventive Maintenance, Burn Prediction, Smart Grid, RUL.**

1. Introduction

Machine maintenance plays a crucial role in the realm of industrial manufacturing and distribution. It contributes to the efficiency, safety, and profitability of manufacturing and distribution operations [6]. On one side, it serves to enhance the operational uptime of machines, thereby minimizing production downtimes. On the other, it also upholds product quality across production lines, reducing defective items, and cost savings in manufacturing and distribution.

In the domain of Smart cities, smart energy is a vital component. It manages the use of technology to improve sustainability, efficiency, and reliability of the energy system. One of the technologies related to this component is Smart grids. A Smart Grid implements some sensors and communication devices to monitor and manage the flow of electricity along a grid, which helpsto manage energy efficiency, reduce energy costs, and increase the reliability of the system.

Transformers play a vital role in a Smart Grid. They perform crucial functions of adjusting voltagelevels for efficient electricity transmission and distribution. Within this context, smart transformers stand out by incorporating sensors and communication technologies, enabling them to actively oversee and manage electricity flow in real time. This data is invaluable to enhance the efficiencyand reliability of the smart grid infrastructure.

Power distribution transformers are critical components within a power distribution system because they are the most expensive elements within a power substation. They are indispensableto the functioning of an electrical power grid as they play a vital role in regulating voltage levels across the various components of the system, spanning from the power generator to the end consumer.

A transformer can experience a variety of problems in its lifecycle that can cause it to malfunc- tion. Some common areas that affect the functionality of transformers are windings, insulations, bushings, and oil. Factors like Voltage, Temperature, Moisture, and Contamination can affect theseareas over time. For example, Windings made of copper or aluminium are prone to overvoltage, overheating, and moisture, Oil serves as a coolant and insulator subject to oxidation, heat, and contamination. Numerous approaches have been suggested and can be broadly categorized into two primary categories: corrective maintenance and preventive maintenance (PdM)[16].

In the context of corrective maintenance, a temporary standby transformer is commonly utilized during repairs conducted at the manufacturer's facility. Nevertheless, the transportation and on- site wiring of these standby transformers following a fault leads to considerable downtime and adversely impacts the System Average Interruption Duration Index (SAIDI)[17].

Hence the importance of preventive maintenance is significant since it performs the maintenancetask at predetermined intervals to stop these faults from occurring. One of the most crucial concepts in many preventive maintenance solutions is to determine its Remaining Useful Life (RUL). RUL refers to the estimated amount of operational time or usage a machine or piece of equipment has left before it is likely to fail or require significant maintenance.

2. Related Works

Several strategies for developing predictive maintenance on transformers have been published in the literature. Some methods use a machine-learning approach while others use a deep learning approach with Multi-sensor Machine-Learning Approaches (MMLA) for classifying multi-sensory data.

The use of Deep Learning model is a novel approach that many papers have tried[1], [15], [3]. The use of BERT[1] as a method for extracting features for long-lived bug prediction is discussed in this article. It analyzes the accuracy of TF-IDF and standard ML classifiers using features obtained with BERT. STAT and FeaR-STAT[15], two newly created transformer-based designs for predicting remaining usable life (RUL) in industrial systems, are presented. The FeaR-STAT model greatly raises the FD004 score when compared to the STAT model, the STAT model offers more uniform predictions. The provided models are useful for RUL estimates and may be used for industrial systems' predictive maintenance, according to the paper's conclusion. This paper[3] suggests a novel model, PM2-CNN for software defect prediction. The model uses a Transformer architecture and multi-channel CNN for classification and combines source code and natural language text to extract semantic features.

Many researchers have also used SVM or models based on SVM in their papers [4], [5], [8], [13], [14]. The method analyzes voltage and current data to identify errors using machine learning methods like Support Vector Machines (SVM), K-Nearest Neighbor (KNN), and Decision Tree (DT)[5]. For a[13] paper, 5 different classifiers were used by training it on a variety of features like Dissolved gas analysis (DGA) results, Physicochemical properties of the oil, and Operating history of the transformer. In this paper[14], the researchers used a novel machine learning approach called Twin Support Vector Machines (TWSVMs) in conjunction with the Chemical Reaction Optimization (CRO) algorithm to develop a predictive model for detecting malicious Domain Generation Algorithms (DGAs). In the results, the average accuracy achieved by the CRO-TWSVM model is 92.15% for all CV folds. When tested without CV the accuracy increased to 94.46%. From this, we can conclude that the CRO-TWSVM model performs effectively while detecting malicious DGAs.

In this paper[8], to train the model, the researchers used real-world data from Colombia's Cauca Department. This data included information on transformer's age, condition, and operating history. The researchers also used two machine learning algorithms: Support Vector Machines (SVMs) and k-means clustering. The researchers used SVMs to learn the features that are most predictive of transformer failure. They then used k-means clustering to group the transformers into different risk categories. The test set contained transformers that were not used to train the model. The researchers found that the model was able to predict transformer failure with an accuracy of 96.74% for 2019 and 97.25% for 2020.

In the area of smart grids, where new algorithms like extreme learning machines (ELM) can play a critical role in predicting and Identifying transformer breakdowns, this move toward predictive maintenance is viewed as important to improving the safety and security of power systems.

3. Methodology

The presented approach makes use of a distribution transformer dataset that was obtained from the Cauca Department of the University of Cauca in Colombia. This department has made available a dataset that includes around 16,000 samples that were obtained from distribution transformers located in 42 municipalities in various urban and rural regions of Colombia. The Cauca Department gathered the data in 2019 and 2020 in an organized manner. The distribution transformers in the dataset were linked to the operator's grid and operated at 13.2 kV and 34.5kV voltages, respectively. There are six categorical variables and five continuous variables in the dataset. Excel spreadsheets organized by the years in which the data was collected. This means that these files must be merged in order to construct the model.

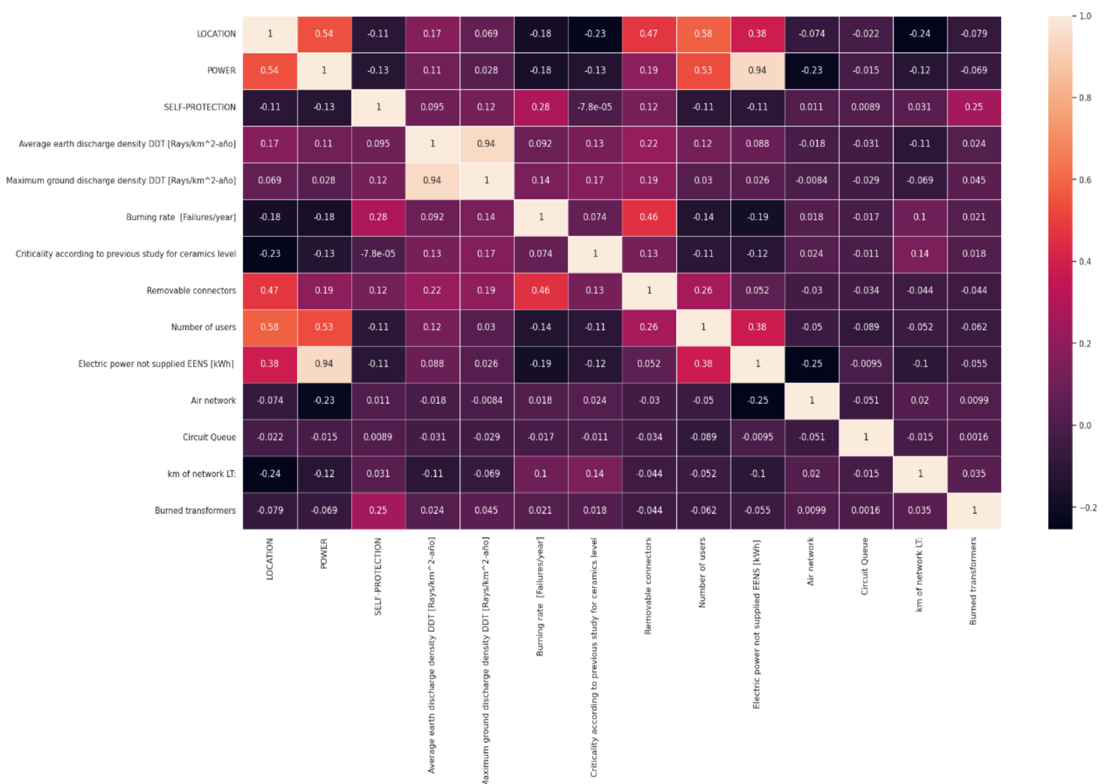

Figure 1 Correlation Heat map

In figure 1 the correlation matrix of the dataset can be seen by the heatmap, which makes correlations between variables clear. A higher electric power unavailable during failures is appears to be correlated with better transformer capacity, for instance, by a positive correlation discovered between "Power" and "Electric power not supplied (EENS)." Also, a possible inverse correlation between "Burned transformers" and "Self-protection" may indicate that internal low voltage protection reduces the possibility of burn events occurring in transformers. It becomes easier to find key connections in the dataset by using this graphical form.

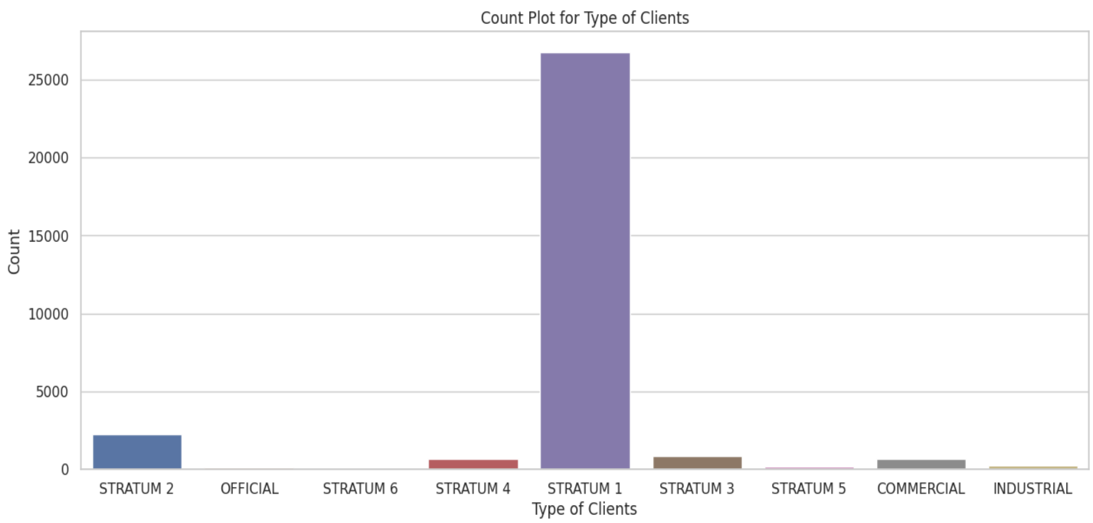

Figure 2 Count plot types of Clients

In Figure 2, The 'Type of clients' count plot effectively displays the frequency of transformers serving different client organizations. It provides an overview of their distribution and indicates the most common clientele, which might be residential, commercial, industrial, or something else entirely.

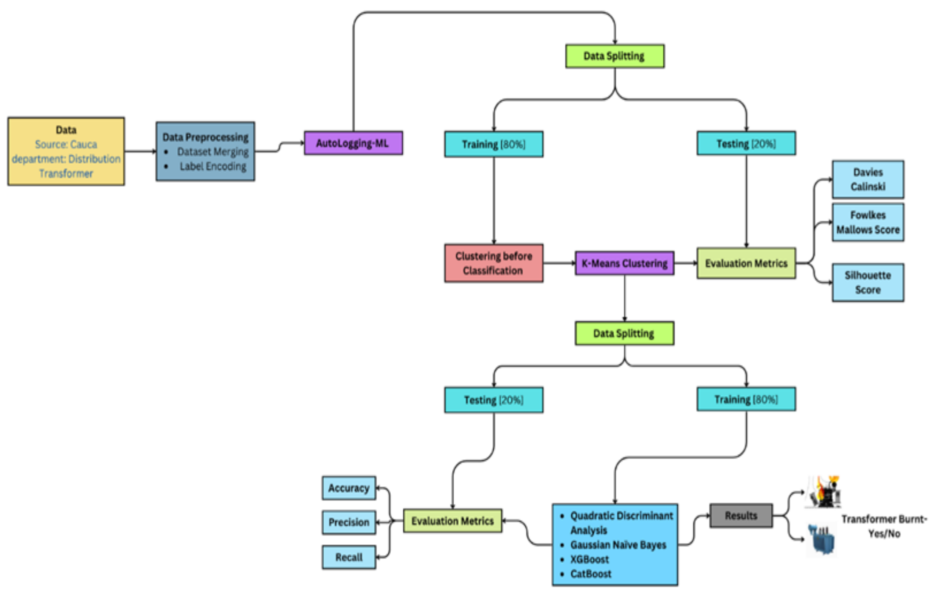

Figure 3 Architecture workflow

Figure 3 gives a brief idea about the proposed model architecture flow that implies after preprocessing the data, a clustering method was used to reduce class imbalance. The modified labels were then given to classification algorithms to anticipate burned transformers, maintaining an extensive and unbiased forecasting process

AutoLogging-ML 1.2.4's classification phase entails training several classifiers on the chosen datasets and then logging performance metrics like accuracy, precision, recall, and F1 score.

About sixteen different classification algorithms are carried out by a specialized function that controls this process. The results, which include accuracy, precision, recall, and F1 score, are diligently entered into an organized table after the training and assessment phase is over which can be seen in the table 1.

Table 1: Autologger Evaluation

ML Models	Training accuracy	Training precision	Training recall	Training f1 score
Quadratic Discriminant Analysis (QDA)	10.9072	52.0607	52.8866	10.8735
Naive Bayes	24.3778	52.2503	58.4073	22.4649
XGboost	97.2003	95.1970	70.5109	77.8220
Catboost Classifier	96.8223	93.6313	66.5821	73.5285

As you can see in above table 1 we have used 4 algorithms among those 16 for the prediction of burned transformers which are Quadratic Discriminant Analysis (QDA), Naive Bayes, XGboost, and Catboost Classifier. As described in table 1 the training accuracy for quadratic discriminant analysis and naive Bayes is relatively very low as compared to the XGboost and Catboost. Regardless of this besides the training accuracy, the precision, recall and f1 scores are also very low for all the algorithms this could happen due to the class imbalance in the labels.The data in the labels are biased towards not burned with 30310 samples and for burned it isonly for 1436 samples. So the reduction in precision, recall, and f1 score can be due to this class imbalance.

Table 1 shows the results of forecasting burned transformers using four algorithms chosen from a pool of 16, namely Quadratic Discriminant Analysis, Naive Bayes, XGBoost, and CatBoost Classifier. Table 1 shows that the training accuracy for Quadratic Discriminant Analysis and Naive Bayes is significantly lower than for XGBoost and CatBoost. Beyond training accuracy,the precision, recall, and F1 score metrics all show a substantial decline among all algorithms. That identified drop in metrics of performance could be related to the labeled data’s inherent classimbalance. The dataset has been biased towards the ’not burned’ class, with 30,310 instances, whereas the ’burned’ class has only 1,436 instances. The outcomes of this class imbalanceare visible in the lower precision, recall, and F1 score values, illustrating the impact of the biased distribution of classes on the entire evolution metrics of the chosen algorithms.

To deal with the problem of class inequalities in labeled data, we used a clustering processthat included the k-means algorithm. This proactive approach includes clustering to generate newlabels, which are then given to the Quadratic Discriminant Analysis, Naive Bayes, XGBoost, and CatBoost Classifier models for classification. This approach aims to reduce the adverse impacts of class inequalities on classification predictions through the use of clustering-generated labels.

4. Result

For the evaluation of clustering the following measures are used Fowlkes-Mallows Score, Silhouette Score, Davies-Bouldin Index and Calinski-Harabasz Index. Their results can be seen in Table 2. The Fowlkes-Mallows Score is a metric for assessing clustering accuracy by comparing true and predicted clusters. It takes the precision and recall of the clustering results into account and returns a single score ranging from 0 to 1, with 1 indicating perfect clustering.

Table 2: Clustering Evaluation

Fowlkes-Mallows	0.7576
Silhouette	0.9061
Davies	0.2395
Calinski	257737.2411

The Silhouette Score compares the resemblance a component is to its own cluster to other clusters. It measures the extent of clustering and ranges from -1 to 1. A larger Silhouette Score signifies that the clusters are more tightly knit and distinct from each other.

The Davies-Bouldin Index measures cluster density and division. It is calculated by taking the mean distance among points within various clusters and dividing it by the mean dimension of the clusters. A smaller Davies-Bouldin Index indicates more distinct clusters.

The Calinski-Harabasz Index, formerly referred to as the Variance Ratio Standard is a cluster validity standard that takes into account the ratio of between-cluster variance to within-cluster variance. This index's larger numbers signify better-defined clusters.

Table 3: Before Clustering Evaluation

Before Clustering				
ML Models	**Training Accuracy (%)**	**Training Precision (%)**	**Training Recall (%)**	**Training F1 score (%)**
Quadratic Discriminant Analysis (QDA)	10.9072	52.0607	52.8866	10.8735
Naive Bayes	24.3778	52.2503	58.4073	22.4649
XGboost	97.2003	95.1970	70.5109	77.8220
Catboost Classifier	96.8223	93.6313	66.5821	73.5285

Table 4: After Clustering Evaluation

After Clustering				
ML Models	**Training Accuracy (%)**	**Training Precision (%)**	**Training Recall (%)**	**Training F1 score (%)**
Quadratic Discriminant Analysis (QDA)	98.1574	98.1574	98.1574	98.1574
Naive Bayes	98.1574	98.1574	98.1574	98.1574
XGboost	99.8897	99.8897	99.8897	99.8897
Catboost Classifier	99.7795	99.7795	99.7795	99.7795

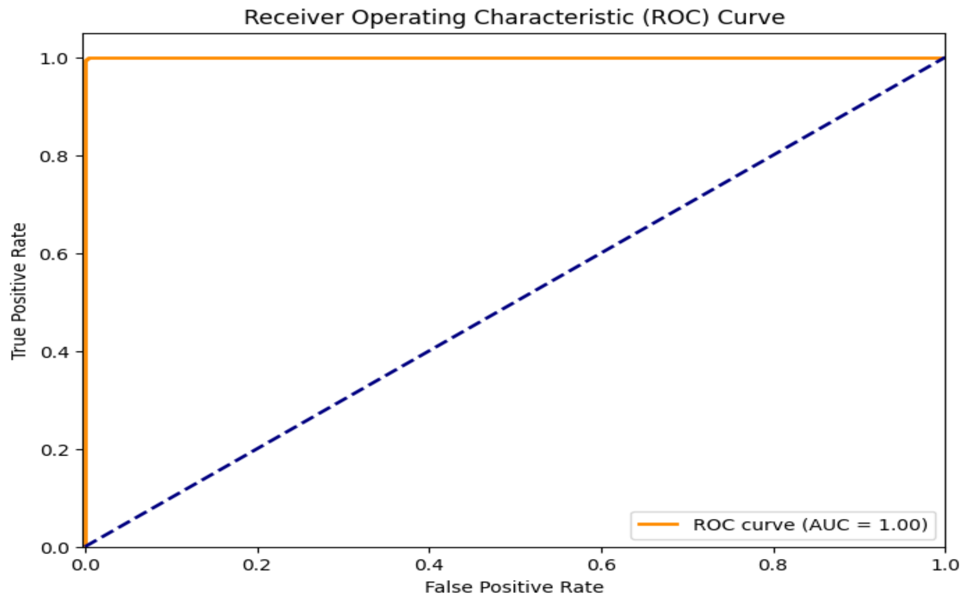

Figure 4 ROC Curve

A binary classifier's performance, like a K-Means clustering result, is graphically measured by the ROC (Receiver Operating Characteristic) curve, which demonstrates the trade-off between true positive rate and false positive rate across different classification thresholds. Better model discriminative capability is can be seen by a greater area under the ROC curve; a curve that approaches the top-left corner indicates stronger performance in identifying between the two clusters.

4.0.1. Front-end Deployment

In crafting the Graphical User Interface (GUI) for deployment, we employed the Streamlit framework. Streamlit, an open-source platform, utilizes the Python language and is specifically designed for constructing web applications geared towards machine learning and data science. Importantly, it exhibits seamless compatibility with key Python libraries, including but not limited to Scikit-learn, Keras, PyTorch, SymPy (latex), NumPy, pandas, and Matplotlib.

In the realm of Graphical User Interface (GUI) deployment for machine learning models, a common practice involves storing model files in the ".pkl" format using the pickle library. These files are subsequently loaded and integrated into Streamlit, facilitating seamless interaction with the model. Notably, the employed model leverages the XGBoost algorithm for predictive analytics. It is imperative to highlight that the preprocessing steps, including label encoding for categorical data and dataset scalarization, are not encapsulated within the model file. Instead, these preprocessing tasks are implemented directly in the GUI code, addressing object data types in two specific columns—namely, "Types of Client" and "Types of Installation."

The GUI development extends beyond model integration, incorporating HTML and CSS to craft dynamic and interactive web pages. Within this framework, a data entry component is strategically designed to align with the model's input requirements. The user is presented with a visually appealing and user-friendly interface, wherein data can be inputted in accordance with the model's specifications.

In figure 5, we have introduced the Main Page for over front-end. Here the user can input values for the various parameters of the transformers.

This approach not only ensures the seamless execution of label encoding for categorical variables but also encompasses scalarization for the dataset, enhancing the model's predictive capabilities. The integration of HTML and CSS elements further contributes to the overall user experience, transforming the GUI into an intuitive platform for data entry and model interaction.

In figure 6, the model takes these inputs and tries to predict the state of the machine. The model gave a positive response for whether it is burned or not. Consequently, in figure 7, the model gives a negative response for a different set of inputs.

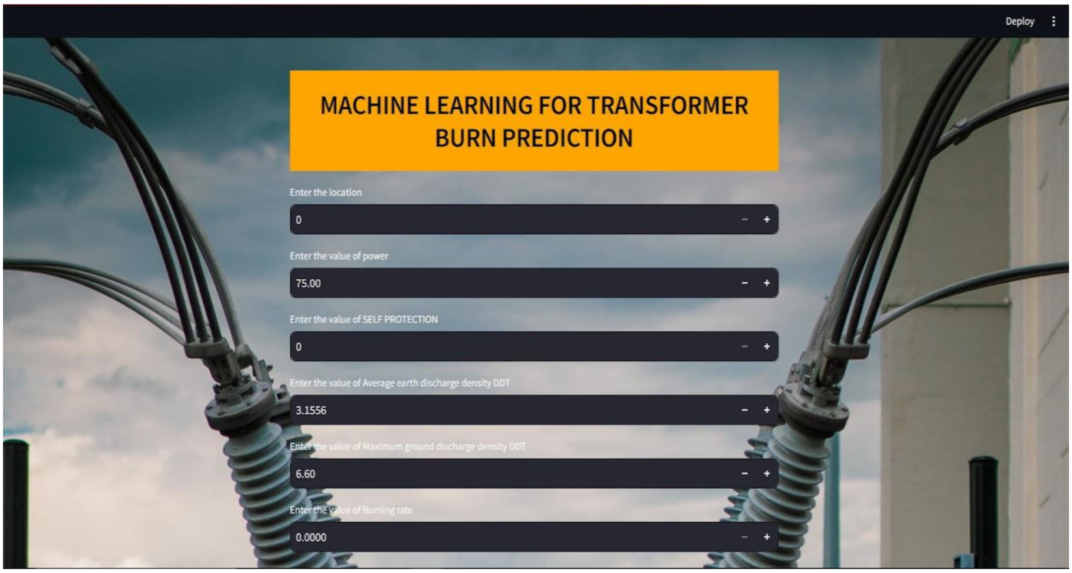

Figure 5 Front-end Main page

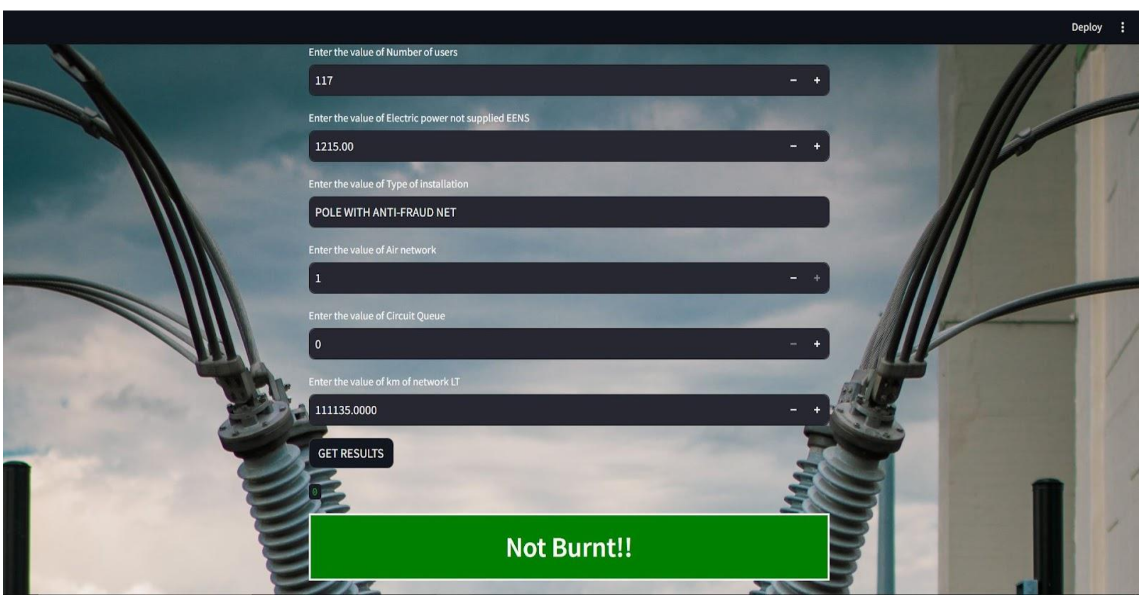

Figure 6 GUI with Positive Burnt Prediction

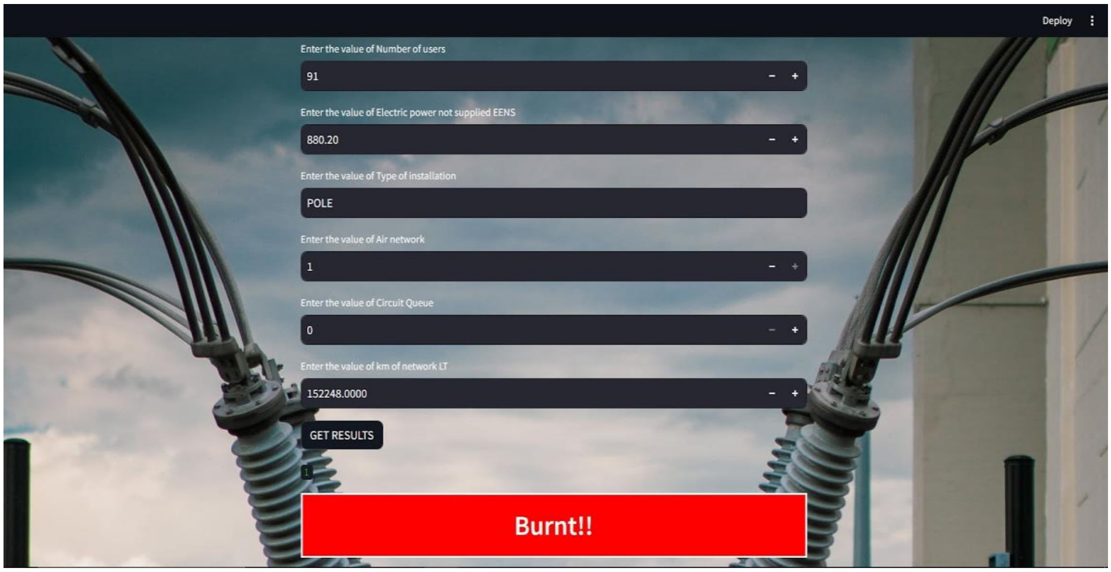

Figure 7 GUI with Negative Burnt Prediction

5. Conclusions

Hence, we have created a pipeline using clustering followed by 4 different classification models that provide as with high accuracy while predicting burn damage for the transformer. The use of clustering was to handle the class imbalance in the data and different models were used to gaugethe performance of the class on different parameters. Despite the high accuracy, the challenges with predicting are still ever present. The limitations stem from the fact that the number of factorsthat affect the working of the transformer is numerous. Some of the external factors are even more difficult to predict than the fault prediction itself like temperature, humidity, voltage surge, and Natural Disasters. Still, the use of such failure prediction models is paramount in reducing time and money. Overall, the main objective is always the reduction of downtime of the machine and to prevent further damage to various other components when one of them fails.

5. Future Scope

Despite the many challenges for fault prediction, the practical effects of this system cannot be discounted. A real-time implementation of such a system that uses real-time sensors to take the input and provides constant updates for the health of the system can help save a lot of time and resources. Such an implementation when used on smart systems like in smart cities where the automation of various systems can significantly increase the efficiency and robustness of many such prediction models.

References

[1]Luiz Gomes, Ricardo da Silva Torres, Mario Lúcio Côrtes,BERT- and TF-IDF-based feature extraction for long-lived bug prediction in FLOSS: A comparative study, Information and Software Technology, Volume 160, 2023,107217, ISSN 0950-5849, https://doi.org/10.1016/j.infsof.2023.107217

[2]Qian Wu, Xiaoyi Yang, and Renming Deng, Predictive maintenance strategy of running fault based on ELM algorithm for power transformer, International Journal of Internet Manufacturing and Services 2018 5:2-3, 297-309, https://doi.org/10.1504/IJIMS.2018.091999

[3]Jingyu Liu, Jun Ai, Minyan Lu, Jie Wang, Haoxiang Shi, Semantic feature learning for software defect prediction from source code and external knowledge, Journal of Systems and Software, Volume 204, 2023, 111753, ISSN 0164-1212, https://doi.org/10.1016/j.jss.2023.111753.

[4]Sheng-wei Fei, Xiao-bin Zhang, Fault diagnosis of power transformer based on support vector machine with genetic algorithm, Expert Systems with Applications, Volume 36, Issue 8, 2009, Pages 11352-11357, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2009.03.022.

[5]Alvarez Quiñones, L. I., Lozano-Moncada, C. A., & Bravo Montenegro, D. A. (2023). Machine learning for predictive maintenance scheduling of distribution transformers. Journal of Quality in Maintenance Engineering, 29(1), 188-202.

[6]Mogos, A. S., Liang, X., & Chung, C. Y. (2023). Distribution Transformer Failure Prediction for Predictive Maintenance Using Hybrid One-Class Deep SVDD Classification and Lightning Strike Failures Data. IEEE Transactions on Power Delivery.

[7]Arias, P. A. (2018). Planning models for distribution grid: A brief review. U. Porto Journal of Engineering, 4(1), 42-55.

[8]Vita, V., Fotis, G., Chobanov, V., Pavlatos, C., & Mladenov, V. (2023). Predictive Maintenance for Distribution System Operators in Increasing Transformers' Reliability. Electronics, 12(6), 1356.

[9]R. A. Sowah et al., "Design of Power Distribution Network Fault Data Collector for Fault Detection, Location and Classification using Machine Learning," 2018 IEEE 7th International Conference on Adaptive Science & Technology (ICAST), Accra, Ghana, 2018, pp. 1-8, doi: 10.1109/ICASTECH.2018.8506774.

[10]Nguyen Huu Du, Nguyen Hoang Long, Kieu Ngan Ha, Nguyen Viet Hoang, Truong Thu Huong, Kim Phuc Tran, Trans-Lighter: A light-weight federated learning-based architecture for Remaining Useful Lifetime prediction, Computers in Industry, Volume 148, 2023, 103888, ISSN 0166-3615, https://doi.org/10.1016/j.compind.2023.103888.

[11] Badawi, M., Ibrahim, S. A., Mansour, D. E. A., El-Faraskoury, A. A., Ward, S. A., Mahmoud, K., ... & Darwish, M. M. (2022). Reliable estimation for health index of transformer oil based on novel combined predictive maintenance techniques. IEEE Access, 10, 25954-25972.

[12] Benmahamed, Y., Kherif, O., Teguar, M., Boubakeur, A., & Ghoneim, S. S. (2021). Accuracy improvement of transformer faults diagnostic based on DGA data using SVM-BA classifier. Energies, 14(10), 2970.

[13] Alqudsi, A., & El-Hag, A. (2019). Application of machine learning in transformer health index prediction. Energies, 12(14), 2694.

[14Yuan, F., Guo, J., Xiao, Z., Zeng, B., Zhu, W., & Huang, S. (2019). A transformer fault diagnosis model based on chemical reaction optimization and twin support vector machine. Energies, 12(5), 960.]

[15]G. S. Chadha, S. R. B. Shah, A. Schwung and S. X. Ding, "Shared Temporal Attention Transformer for Remaining Useful Lifetime Estimation," in IEEE Access, vol. 10, pp. 74244-74258, 2022, doi: 10.1109/ACCESS.2022.3187702.

[16]EN, C. (2001). 13306: Maintenance terminology. European Committee for Standardization: Brussels, Belgium.

[17]Ghosh, S., & Dutta, S. (2021). Ensemble machine learning methods for better dynamic assessment of transformer status. Journal of The Institution of Engineers (India): Series B, 102(5), 1113-1122.