Hindawi

Research Article

Analysis of Noise Pollution during Dussehra Festival in Bhubaneswar Smart City in India: A Study Using Machine Intelligence Models

Sourav Kumar Bhoi,[1] Chittaranjan Mallick,[2] Chitta Ranjan Mohanty,[3] and Ranjan Soumya Nayak [4]

[1]*Department of Computer Science and Engineering, Parala Maharaja Engineering College (Govt.), Brahmapur 761003, Odisha, India*
[2]*Department of Basic Science (Mathematics), Parala Maharaja Engineering College (Govt.), Brahmapur 761003, Odisha, India*
[3]*Department of Civil Engineering, Parala Maharaja Engineering College (Govt.), Brahmapur 761003, Odisha, India*
[4]*Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India*

Correspondence should be addressed to Ranjan Soumya Nayak; srnayak@amity.edu

Controlling noise pollution in smart cities is a big challenge nowadays due to rise in urbanization and industrialization. As population mass grows, the celebration of yearly festivals such as Dussehra in Bhubaneswar city is also getting popular. However, since this sound pollution is creating a risk to human health, regular monitoring is strictly needed. In this work, the noise pollution level of Bhubaneswar smart city during Dussehra 2020 is predicted using different supervised machine learning (ML) prediction models. The input parameters considered for this work are area or zones of Bhubaneswar city, time at which sound level recorded, equivalent continuous sound level (Leq in dBA), and noise level (high/low compared to the standard value). The data collected for training phase and testing phase by using different ML models is taken from State Pollution Control Board, Odisha, India, for the years 2015–2020. The supervised ML models taken in this work are Decision Tree (DT), Neural Network (NN), k-Nearest Neighbor (k-NN), Naïve Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF). The predictions of the models are evaluated using Orange 3.26 data analytics tool. From the results, it was found that DT and RF show a higher classification accuracy, 92.5%, than that of other ML models. Moreover, it is observed that the probability of prediction of noise pollution level for the testing dataset for DT is higher for high noise level and for RF is higher for low noise level than other prediction models.

1. Introduction

Bhubaneswar city has an ancient history of 3000 years, being founded during the Kalinga Empire [1]. The modern Bhubaneswar city was established in 1948 in Odisha state of India as shown in Figure 1. It is also known as the temple city of Odisha. It has many munificent temples. The modern Bhubaneswar was planned and designed by Otto Konigsberger in 1946, and the population of the city is 8.41 lakh as per 2011 census [1]. This city is also well known for hubs such as education, software and IT industry, tourism, industry and factories, and hospitals. It has good socio-economic balance and cultural development. However, the urbanization and industrialization in the city lead to many types of pollution. One of the big challenging problems is noise pollution [3, 4]. The main sources of noise pollution are traffic noise, industry or factories noise, house noise, animals' noise, people noise, speakers, air traffic noise, construction noise, hotels or restaurants noise, machine noise, natural calamity, accidents noise, festivals noise, crackers noise, etc. This causes many disorders to the life such as physical disorders, psychological disorders, and

behavioural disorders. Therefore, the noise pollution should be detected and controlled well in the city before it reaches a level where it affects humans and animals.

Bhubaneswar is well known for Dussehra festival or Vijayadashami where people from different parts of the country as well as from parts of Odisha state comes to celebrate. At many locations, streets, or zones, Dussehra is celebrated for 10 days. On these 10 days, there are huge gatherings at the zones where idols are worshipped. This creates huge noise due to gatherings, traffic, speakers, celebration, etc. During the time of Ravana burning ceremony, a lot of cracker's sounds, traffic sounds, etc. also occur. At the time of Durga immersion, there are processions which create noisy environments by speakers, traffic, gatherings, etc. Therefore, proper controlling and management of noise pollution during Dussehra in Bhubaneswar should be done every day so that the noise level becomes below the standards in specific regions [5] such as commercial region, silence region, industrial region, and residential region. The areas considered in this work are shown in Figure 2.

Machine learning nowadays is a booming area of research in the field of Artificial Intelligence, where the ML models take data as input, process it, and predict the output [7–9]. ML consists of many methods for data analysis such as supervised learning models, unsupervised learning models, reinforcement models, and semi-supervised models [7]. ML has many applications in healthcare, industry, transportation, organizations, agriculture, technology, etc. ML will be best suited to the above monitoring problem [10, 11] where the noise level is estimated based on the previous features. In this study, a prediction of noise levels in Bhubaneswar during Dussehra 2020 is made using supervised ML models. Due to their high accuracy and suitability for our classification problem of predicting noise level as either high or low, the most widely accepted models in the field of supervised machine learning (ML), including the DT, NN, k-NN, NV, and RF, are applied. They work well in smaller datasets. In this work, the dataset size is smaller and can be well handled by the supervised ML algorithms. However, if the dataset size also increases, the models can be updated, which helps in increasing the prediction accuracy.

Many such research works have been made in the area of noise pollution monitoring in major cities of Odisha, India. Noise assessments are made using different noise models at Urban Parks of Bhubaneswar and Puri, NH 316, Rourkela city, Indian offices, and Bhubaneswar city [12–16]. Swain et al. [12] performed a case study at 10 different office corridors in the city of Balasore. The noise levels are monitored at different timings such as 10–12 pm, 1–3 pm, and 3–5 pm. At RTO, the recorded maximum sound level was 83.4 dB. Swain et al. [13] developed several models to predict the noise level at NH 316 using different noise descriptors. The experimental data is collected from seven different squares. Swain et al. [14] analyzed the noise levels at 16 different squares of Bhubaneswar city to monitor the road traffic noise using several noise descriptors. Swain et al. [15] analyzed noise levels at 3 parks in Bhubaneswar and Puri. A

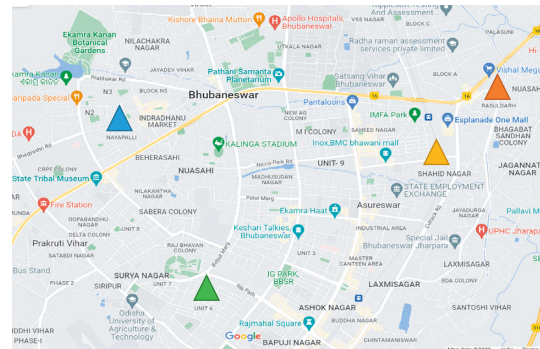Figure 1: Google map showing Bhubaneswar smart city (red icon) of Odisha state in India [2].

Figure 2: Google map showing different areas or zones of Bhubaneswar (latitude: 20.296059 and longitude: 85.824539) where the noise data is considered [6]. Blue triangle: Nayapalli area, residential zone; green triangle: Capital Hospital area, silence zone; yellow triangle: Shahid Nagar area, commercial zone; orange triangle: Rasulgarh area, industrial zone.

questionnaire was prepared, and 330 participants joined it to check the quality of the parks with respect to sound. Goswami et al. [16] performed a study at 12 different squares in the city of Rourkela. The noise levels are monitored at different timings between 7 am to next day 6 am. In Berhampur city and Sambalpur city of Odisha, India, many assessments were also done for noise pollution for road traffic, festivals, etc. [17–20]. Sahu et al. [17] performed a study at 12 different locations in the city of Sambalpur. The noise levels are monitored at different timings from 6 am to 6 pm. Sahu et al. [18] studied the noise levels in Berhampur city during Diwali festival. They consider three different areas to collect the noise levels. They also consider the air pollution parameters during Diwali due to pollution from crackers. Sahu et al. [19] evaluated the traffic noise at Burla town using regression equations to monitor its effect on the local people and patients. Sahu et al. [20] studied the traffic noise in Berhampur city by considering 11 locations and evaluated the noise levels using multiple linear regression models.

Many such research works have been done in the area of noise pollution study. However, from the above research, it has been found that much less work has been done in the area of noise pollution monitoring in Bhubaneswar during festivals using ML models. Therefore, in this work, an estimation of noise levels has been made for different noise levels at the time of Dussehra 2020 in Bhubaneswar city.

The main contributions in this work are stated as follows:

(1) The noise pollution level of Bhubaneswar smart city during Dussehra 2020 is predicted using different supervised ML models. The input parameters considered for this work are areas or zones of Bhubaneswar city, time at which sound level recorded, equivalent continuous sound level (Leq in dBA), and noise level (high/low compared to the standard value). The areas considered as per the dataset [5] are silence zone (Capital Hospital area), industrial zone (Rasulgarh area), commercial zone (Saheed Nagar area), and residential zone (Nayapalli area).

(2) The data collected for training phase and testing phase using different ML models is taken from State Pollution Control Board (SPCBO), Odisha, India, in 2015–2020 [5]. The supervised ML models taken in this work are DT, NN, k-NN, NV, SVM, and RF.

(3) The predictions of the models are evaluated using Orange 3.26 data analytics tool [21].

The rest of the paper is organized as follows. Section 2 presents the data and methodology. Section 3 presents the results and discussion where the predictions are made for different noise levels of Dussehra 2020. Section 4 presents the conclusion of the whole work.

2. Data and Methodology

In this section, we describe the data collection phase for training and testing, machine learning phase, and estimation of noise level.

2.1. Data Collection. The dataset consists of two sets, namely, training set and testing set. Training dataset is the set that is used as input to the supervised ML models for training to predict the output (noise levels). Testing dataset is the set that is used for testing the ML models that have been already trained to get the predictions accurately with high probability.

2.2. Training and Testing Data. In this section, we describe the training and testing dataset as follows.

2.2.1. Training Dataset. The training dataset is collected from SPCBO, Odisha, India, from Environment Monitoring Data section [5]. The data consist of year-wise noise levels of Bhubaneswar city at four different areas during Dussehra, in 2015-2019. The latitude and longitude of Bhubaneswar city are 20.296059 and 85.824539, respectively. The four areas for which the data are taken are Nayapalli (residential zone), Saheed Nagar (commercial zone), Capital Hospital (silence zone), and Rasulgarh (Industrial zone). The dataset mainly consists of four features as mentioned in Table 1, namely, area, time, Leq, and noise level. Leq is the continuous sound level in decibels. Noise level is determined as high or low with respect to the noise standards at different zones as shown in Table 2. For training, these four features are added to the ML algorithms by assigning noise level as target level. If the noise level is greater than the standard at that area, the noise is considered high; otherwise, it is low.

The histogram of all features with respect to noise is shown in Figure 3. From the figure, it is observed that Leq has a center at 64.072, dispersion of 0.322, minimum value of 0, and maximum value of 95.2, with no missing values. Noise feature center is high, dispersion is 0.687, and minimum and maximum values are not applicable, with no missing values. Time feature center is day, dispersion is 0.637, and minimum and maximum values are not applicable, with no missing values. Area feature center is commercial, dispersion is 1.39, and minimum and maximum values are not applicable, with no missing values. The results are generated using Orange machine analytics tool.

In addition, the distribution of dataset is presented in Figures 4–7. Here, mainly, the frequency of the features is considered. Figure 4 shows the distribution of noise feature with respect to noise feature with 90 high values, 24 low values, and 6 na values. Figure 5 shows the distribution of area feature with respect to noise feature. In commercial area, there are 30 values: 25 high, 3 low, and 2 na. In industrial area, there are 30 values: 13 high, 15 low, and 2 na. In residential area, there are 30 values: 28 high, 2 low, and 0 na. In silence area, there are 30 values: 24 high, 4 low, and 2 na. Figure 6 shows the distribution of time with respect to noise. Day has 80 values: 57 high, 17 low, and 6 na. Night has 40 values: 33 high, 7 low, and 0 na. Figure 7 shows frequency vs. Leq graph where the distribution of Leq feature with respect to noise is represented.

For training, we have considered 2015–2019 noise pollution data at the time of Dussehra. Therefore, the number of instances is $24 \times 5 = 120$ instances (rows), where 24 is the number of instances in a year as per Table 3 and 5 is the number of years (2015–2019).

2.2.2. Testing Dataset. For testing, we have collected data for Dussehra in 2020 from State Pollution Control Board, Odisha, which was recorded on 26.10.2020. This is the time when restrictions were going on due to pandemic. Therefore, there are also 24 instances for year 2020. By the help of trained ML prediction models, the noise level (high/low) for each instance is predicted.

2.3. Machine Learning and Prediction of Noise Level. In this work, we have considered 6 supervised machine learning prediction models for predicting the output (noise level) for the testing dataset. The models considered are NN, k-NN, RF, DT, NB, and SVM [7–9]. These models work best with this type of classification problems where the class is predicted based on high or low noise levels. These models take the input as training set and divide the set using sampling technique (k-fold) for training the models. The model with high accuracy is considered as the model that can be used for prediction of the noise levels. However, in this work, all models are trained and all models are tested for prediction of the noise level.

TABLE 1: Sample average data of year 2017 during Dussehra.

Area	Time (different hourly timings at day and night)	Average Leq (dBA) in hr	Noise
Residential	Day	71.1	High
Residential	Day	65.3	High
Residential	Day	66.9	High
Residential	Day	63.0	High
Residential	Night	60.8	High
Residential	Night	77.8	High
Commercial	Day	72.8	High
Commercial	Day	74.5	High
Commercial	Day	73.2	High
Commercial	Day	77.0	High
Commercial	Night	68.8	High
Commercial	Night	66.3	High
Silence	Day	60.7	High
Silence	Day	60.8	High
Silence	Day	59.7	High
Silence	Day	61.8	High
Silence	Night	57.2	High
Silence	Night	53.4	High
Industrial	Day	74.7	Low
Industrial	Day	77.4	High
Industrial	Day	71.9	Low
Industrial	Day	71.7	Low
Industrial	Night	70.1	High
Industrial	Night	66.9	Low

TABLE 2: Noise standards as per State Pollution Control Board, Odisha.

Zones	Day/night	Standards
Residential	Day	55
	Night	45
Commercial	Day	65
	Night	55
Silence	Day	50
	Night	40
Industrial	Day	75
	Night	70

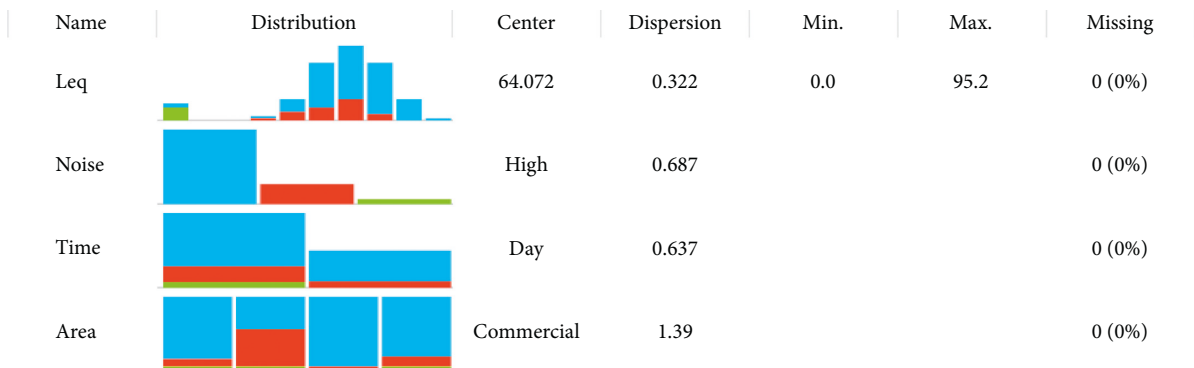| Name	Distribution	Center	Dispersion	Min.	Max.	Missing	
Leq		64.072	0.322	0.0	95.2	0 (0%)	
Noise		High	0.687			0 (0%)	
Time		Day	0.637			0 (0%)	
Area		Commercial	1.39			0 (0%)	

FIGURE 3: Histogram generated using Orange tool for different dataset features with respect to noise feature (blue: high, red: low, green: na).
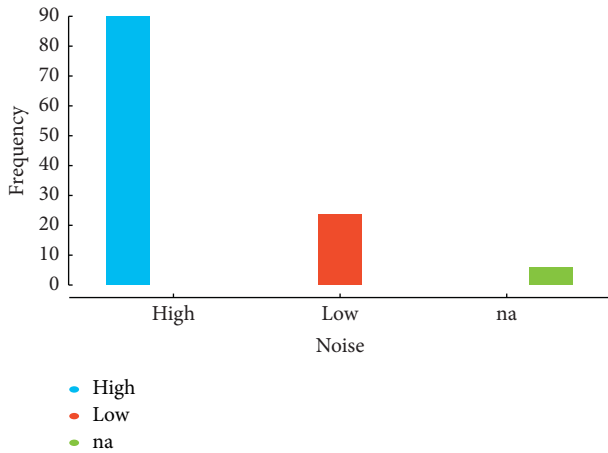
FIGURE 4: Noise feature distribution in dataset generated by Orange tool with respect to noise feature.
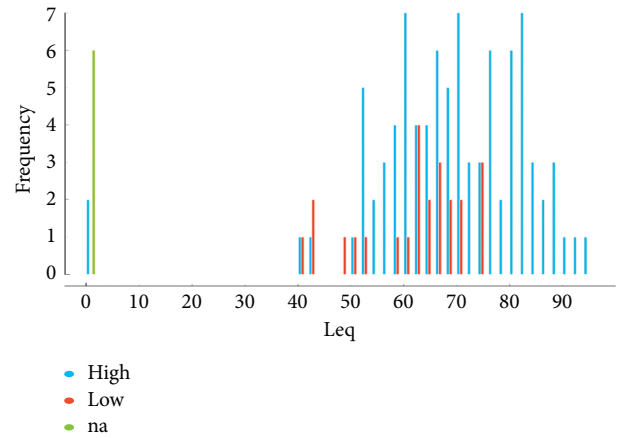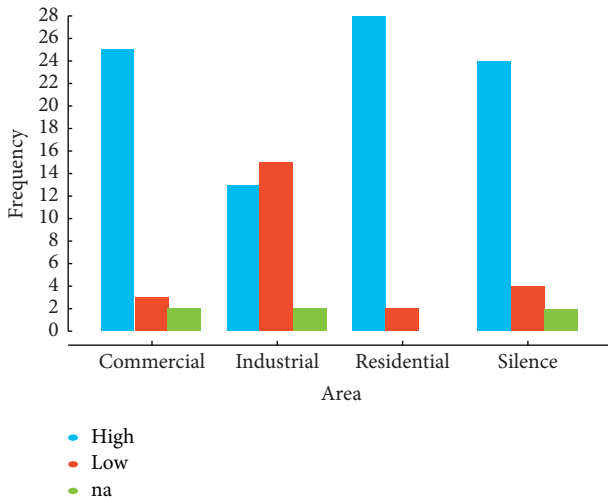

FIGURE 5: Area feature distribution in dataset generated by Orange tool with respect to noise feature.

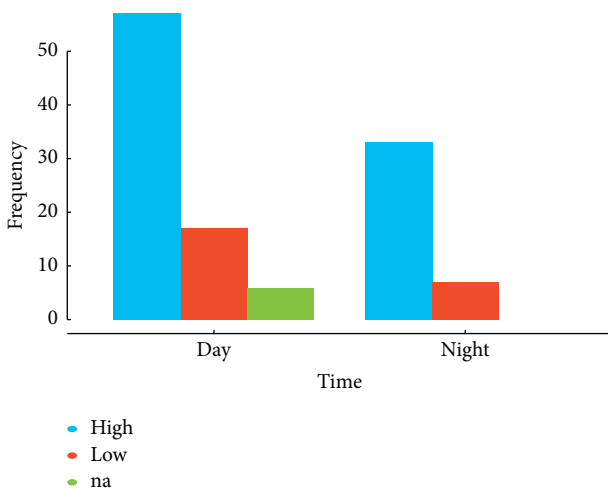FIGURE 6: Time feature distribution in dataset generated by Orange tool with respect to noise feature.

FIGURE 7: Leq feature distribution in dataset generated by Orange tool with respect to noise feature.

TABLE 3: The performance parameters generated using the Orange tool for evaluation.

Models	AUC	CA	F1	Precision	Recall
k-NN	0.849	0.825	0.802	0.814	0.825
DT	0.941	0.925	0.922	0.926	0.925
SVM	0.962	0.858	0.849	0.848	0.858
RF	0.976	0.925	0.926	0.928	0.925
NN	0.960	0.908	0.907	0.909	0.908
Naïve Bayes	0.855	0.858	0.831	0.811	0.858

2.3.1. Steps for Prediction of Noise Level. The steps for predicting the noise level are shown in Figure 8 and discussed as follows:

Step 1. The input is taken as the training dataset and given into the ML model.

Step 2. The ML model uses k-fold as the sampling technique for training and testing.

Step 3. The evaluation results are then analyzed to know the model that has a higher classification accuracy (CA).

Step 4. The model with higher CA is selected for predicting the noise level.

Step 5. The testing is done by taking the testing dataset as input and given into the prediction model to get the target noise level with higher probability (0-1).

Step 6. The prediction model that shows higher probability of estimation of water class for the input parameters is considered for taking the noise level data.

3. Results and Discussion

The performance of the methodology is evaluated using Orange 3.26 data analytics tool [21] installed in a Core i3 machine with 8 GB RAM, 2.4 GHz processor, and 64 bit Windows 10 OS platform. The model has been simulated with different parameter consideration which is properly presented in subsequent section, all the nonlinearity of
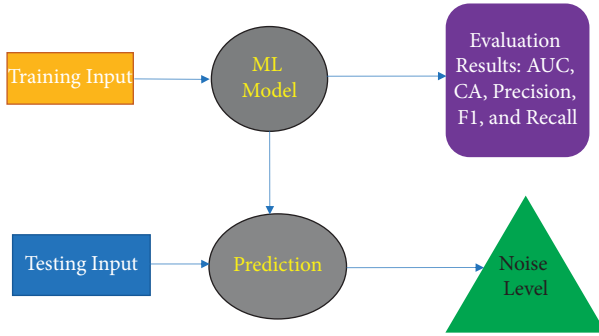
Figure 8: Flowchart of prediction model for estimating the noise level.

model has been achieved by imposing ReLU activation function in case of NN, and sampling technique has been set as k-10. The configuration setting of different ML models is discussed as follows.

k-NN is a supervised machine learning approach that assumes similarity between the new data and available data. Afterward, it puts the new data into the group of the most nearest/similar available data. Firstly, number k is selected for the neighbors, and then Euclidean distance is calculated for k number of neighbors. Then, as per the Euclidean distance, take the k-nearest neighbors. Then, by using the k neighbors, the amount of data in each category is calculated. Then, the new data is assigned to the category that has the maximum number of neighbors. For implementation of k-NN through Orange here, the number of neighbors is considered as 5, the metric taken is Euclidean distance to calculate the distance of the k number of neighbors, and the weight is set to uniform. DT is a supervised machine learning approach mainly used for classification problems. It is a tree structure classifier where the leaf nodes are outcome, internal nodes are the features of the dataset, and branches are the decision rules. It mainly starts from the root node, say R, that contains the dataset; then, it finds the best attribute using the attribute selection measure. Afterward, S is divided into subsets that contain the possible values for the attributes selected. Then, the decision tree node that contains the best attribute is generated. Then, new decision trees are recursively generated by taking the subsets found earlier from S. This process is continued till the nodes are not further classified, and that node is called leaf node. For implementation of DT through Orange here, the binary tree is induced with minimum number of instances in the leaves set to 2, maximum subset splitting is done till 5, maximal depth in tree is 100%, and stopping condition when majority is reached is set to 95%. SVM is mainly used for classification problems by putting a best line or decision boundary to segregate the n dimensional space into different classes. This best line or boundary is called a hyperplane. It mainly chooses the extreme points called support vectors for creating the hyperplane. For implementation of SVM through Orange here, the parameters are set as follows: cost is set to 1.0, regression loss is set to 0.1, kernel is set to RBF, and iteration limit is set to 100. RF is a supervised machine learning approach to solve the classification problem. This

approach is mainly based on ensemble learning where many classifiers are combined to solve the complex classification problem. In RF, firstly, K random points are selected from the dataset, and then decision trees are created based on the selected data points. Then, N number is chosen to create N decision trees, and then the first two steps are repeated. At last, for the new data, predict the outcome of each decision tree and assign the new data to the group that has majority votes. For implementation of RF through Orange here, the number of trees is set to 10 and maximum splitting of subsets is done till 5. NN is a supervised machine learning approach used to solve classification problems. NN is a collection of network functions to take the input and generate desired output. It has three layers: input, hidden, and output layer. The inputs are attached with weights and sent to the neurons at a hidden layer for computing a value. If that value is greater than or less than a threshold value, the decision/output is generated. For implementation of NN through Orange here, the number of neurons is set to 100 with activation function as ReLU, solver is set to Adam, regularization parameter alpha is set to 0.0001, and maximum number of iterations is set to 200. NB is a supervised machine learning approach based on Bayes theorem to solve classification related problems. It is basically a probabilistic classifier that uses the basis of probability of data. In NB, firstly, the dataset is converted into frequency tables, and then the likelihood table is generated by finding the feature probability. Then, Bayes theorem is used to calculate the posterior probability. For implementation of NB through Orange here, the Naïve Bayes Classifier algorithm is used.

The test and score show the different performance parameters of the ML models, and the confusion matrix shows how accurately the instances are predicted from the actual predicted value. The performance parameters taken are:

(1) AUC (area under curve): It describes how much the ML model classifies the classes well. The model with 100% accuracy of prediction has an AUC of 1.0.

(2) CA (classification accuracy): The number of correct predictions made from the observed values is called CA. The following equation shows the formula for CA:

$$CA = \frac{(TP + TN)}{(TP + FP + TN + FN)},\qquad(1)$$

where TP is the true positive, TN is true negative, FP is the false positive, and FN is the false negative.

(3) F1: The harmonic mean of precision and recall to know the accuracy better is shown by F1 score. It is shown in the following equation:

$$F1 = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)}.\qquad(2)$$

(4) Precision: Which instances are accurately classified in a specific positive class as a percentage of all instances of that class that have been classified. The following equation shows the formula for precision:
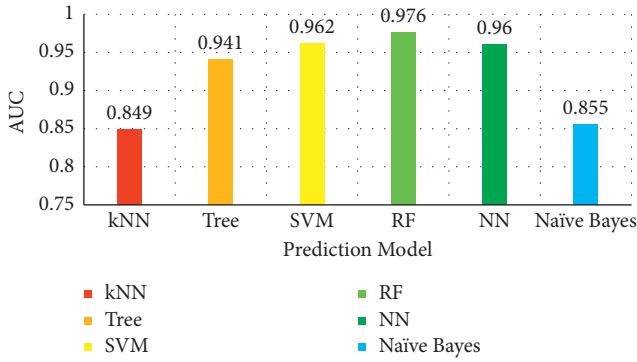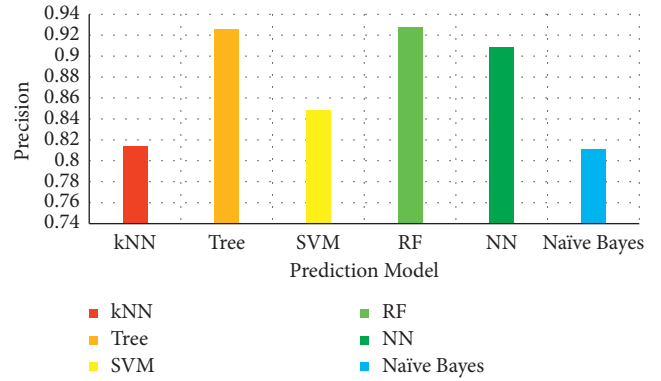
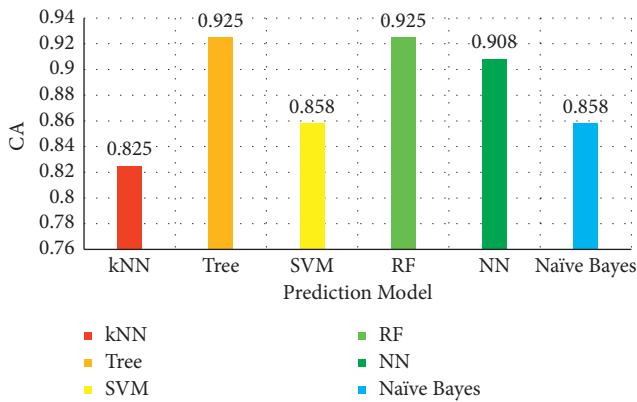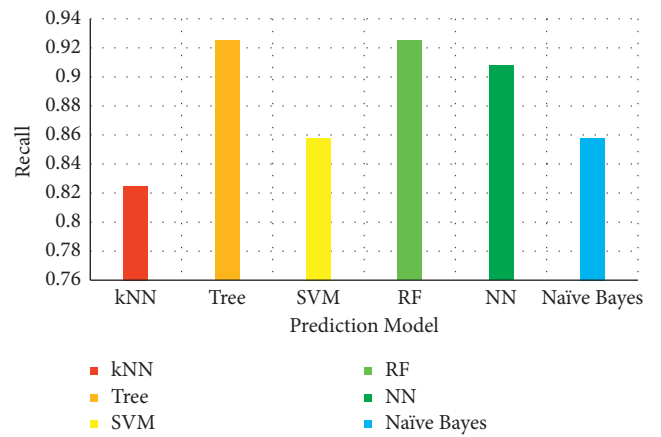Figure 9: The AUC of all ML models.

Figure 10: The CA of all ML models.
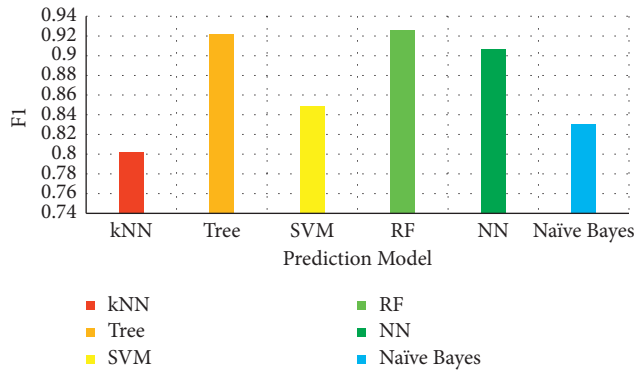

Figure 11: The F1 of all ML models.

$$\text{Precision} = \frac{TP}{(TP + FP)}. \tag{3}$$

(5) Recall: Recall means the proportion of actual instances correctly classified for a particular class. The following equation shows the formula for recall:

$$\text{Recall} = \frac{TP}{(TP + FN)}. \tag{4}$$

From Table 3, it is observed that the CA of DT and RF is greater than that of other ML models. Therefore, we can

Figure 12: The precision of all ML models.

Figure 13: The recall of all ML models.

conclude that this model will be better for prediction. Other performance parameter results are visualized in Figures 9–13. However, we have taken the main parameter as CA for prediction. From Figure 10, it is observed that the AUC values for k-NN, DT, SVM, RF, NN, and NB are 0.849, 0.941, 0.962, 0.976, 0.960, and 0.855, respectively. The AUC for RF is the highest among all the ML models. From Figure 11, it is observed that the CA values for k-NN, DT, SVM, RF, NN, and NB are 0.825, 0.925, 0.858, 0.925, 0.908, and 0.858, respectively. The CA values for DT and RF are the highest among all the ML models. From Figure 12, it is observed that the F1 values for k-NN, DT, SVM, RF, NN, and NB are 0.802, 0.922, 0.849, 0.926, 0.907, and 0.831, respectively. The F1 for RF is the highest among all the ML models. From Figure 13, it is observed that the precision values for k-NN, DT, SVM, RF, NN, and NB are 0.814, 0.926, 0.848, 0.928, 0.909, and 0.811, respectively. The precision for RF is the highest among all the ML models. From Figure 14, it is observed that the recall values for k-NN, DT, SVM, RF, NN, and NB are 0.825, 0.925, 0.858, 0.925, 0.908, and 0.858, respectively. The recall for DT and RF is the highest among all the ML models.

Figure 14 shows the prediction model design using the Orange workflow. Here, each ML model is used for prediction of a particular noise level for each instance (testing set) with a probability value (0.001 to 1.00). Therefore, our main goal is to find the noise level that has higher probability of
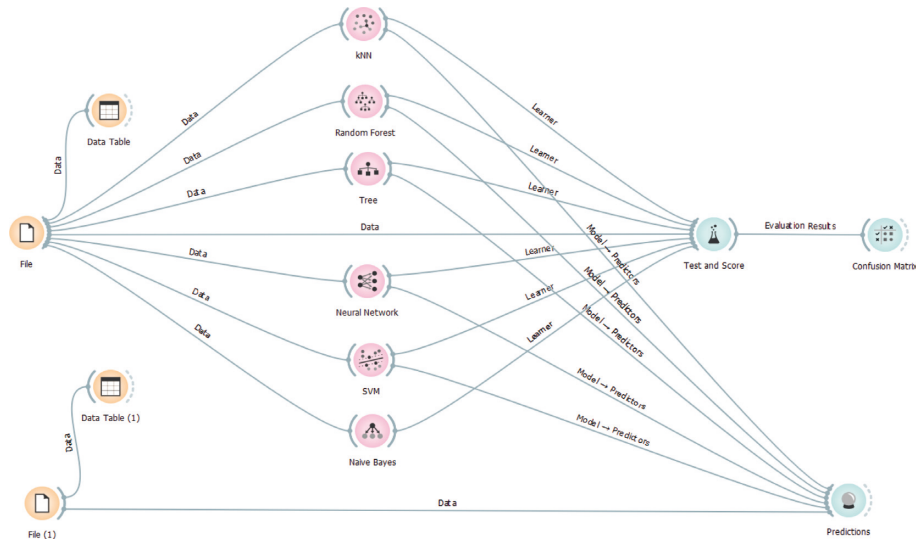
FIGURE 14: Orange tool workflow diagram for evaluating the results of different ML algorithms.

	Tree	Neural Network	SVM	kNN	Naive Bayes	Random Forest	Area	Time	Leq
1	1.00 → High	0.93 → High	0.91 → High	0.80 → High	0.82 → High	0.97 → High	Residential	Day	67.7
2	1.00 → High	0.94 → High	0.92 → High	1.00 → High	0.82 → High	0.97 → High	Residential	Day	70.5
3	1.00 → High	0.92 → High	0.90 → High	0.80 → High	0.82 → High	0.97 → High	Residential	Day	66.8
4	1.00 → High	0.93 → High	0.91 → High	0.80 → High	0.82 → High	0.97 → High	Residential	Day	68.1
5	1.00 → High	0.99 → High	0.90 → High	0.80 → High	0.86 → High	1.00 → High	Residential	Night	64.5
6	1.00 → High	0.98 → High	0.89 → High	0.80 → High	0.85 → High	0.93 → High	Residential	Night	51.7
7	1.00 → High	0.69 → High	0.85 → High	0.80 → High	0.72 → High	1.00 → High	Commercial	Day	66.8
8	1.00 → High	0.90 → High	0.91 → High	0.80 → High	0.72 → High	1.00 → High	Commercial	Day	75.6
9	1.00 → High	0.93 → High	0.93 → High	1.00 → High	0.91 → High	1.00 → High	Commercial	Day	78.4
10	1.00 → High	0.75 → High	0.87 → High	0.80 → High	0.72 → High	1.00 → High	Commercial	Day	68.8
11	1.00 → High	0.98 → High	0.89 → High	0.80 → High	0.72 → High	0.79 → High	Commercial	Night	58.2
12	1.00 → High	0.99 → High	0.91 → High	1.00 → High	0.80 → High	1.00 → High	Commercial	Night	71.2
13	1.00 → High	0.77 → High	0.85 → High	1.00 → High	0.45 → High	0.82 → High	Silence	Day	56.1
14	1.00 → High	0.84 → High	0.88 → High	1.00 → High	0.45 → High	1.00 → High	Silence	Day	59.1
15	1.00 → High	0.74 → High	0.83 → High	1.00 → High	0.45 → High	0.82 → High	Silence	Day	54.8
16	1.00 → High	0.77 → High	0.85 → High	1.00 → High	0.45 → High	0.82 → High	Silence	Day	55.9
17	1.00 → High	0.98 → High	0.90 → High	0.60 → High	0.69 → High	0.95 → High	Silence	Night	48.5
18	1.00 → High	0.98 → High	0.90 → High	0.40 → Low	0.69 → High	0.95 → High	Silence	Night	44.4
19	0.00 → Low	0.17 → Low	0.08 → Low	0.60 → High	0.29 → Low	0.00 → Low	Industrial	Day	61.7
20	0.00 → Low	0.24 → Low	0.16 → Low	0.40 → Low	0.30 → Low	0.00 → Low	Industrial	Day	67.8
21	0.17 → Low	0.33 → Low	0.25 → Low	0.60 → High	0.30 → Low	0.00 → Low	Industrial	Day	71.3
22	1.00 → High	0.44 → Low	0.39 → Low	0.40 → Low	0.30 → Low	0.10 → Low	Industrial	Day	75.1
23	0.17 → Low	0.34 → Low	0.15 → Low	0.60 → High	0.37 → Low	0.52 → High	Industrial	Night	70.9
24	0.00 → Low	0.15 → Low	0.06 → Low	0.80 → High	0.37 → Low	0.12 → Low	Industrial	Night	53.8

FIGURE 15: Prediction of high noise level using different ML models for the input instances with probability values (0.001–1.00).

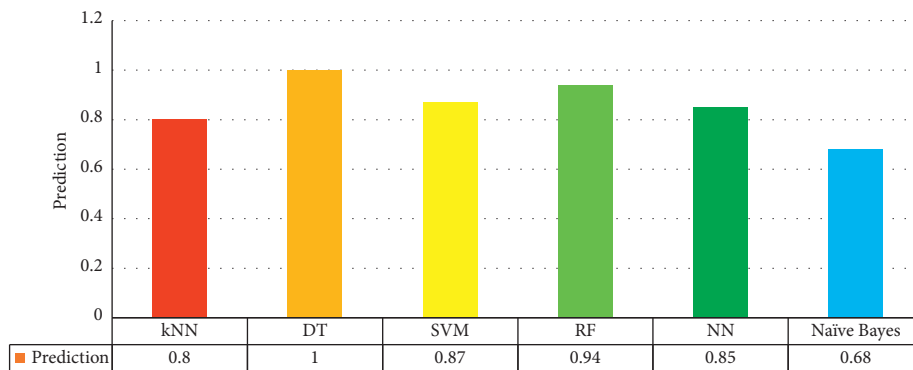|	kNN	DT	SVM	RF	NN	Naïve Bayes
Prediction	0.8	1	0.87	0.94	0.85	0.68

FIGURE 16: Graphical presentation of prediction of high noise level using different ML models for the input instances with probability values (0.001–1.00).

	Tree	Neural Network	SVM	kNN	Naïve Bayes	Random Forest	Area	Time	Leq
1	0.00 → High	0.07 → High	0.09 → High	0.20 → High	0.16 → High	0.03 → High	Residential	Day	67.7
2	0.00 → High	0.05 → High	0.08 → High	0.00 → High	0.16 → High	0.03 → High	Residential	Day	70.5
3	0.00 → High	0.08 → High	0.09 → High	0.20 → High	0.16 → High	0.03 → High	Residential	Day	66.8
4	0.00 → High	0.07 → High	0.09 → High	0.20 → High	0.16 → High	0.03 → High	Residential	Day	68.1
5	0.00 → High	0.01 → High	0.09 → High	0.20 → High	0.13 → High	0.00 → High	Residential	Night	64.5
6	0.00 → High	0.02 → High	0.10 → High	0.20 → High	0.10 → High	0.07 → High	Residential	Night	51.7
7	0.00 → High	0.28 → High	0.15 → High	0.20 → High	0.21 → High	0.00 → High	Commercial	Day	66.8
8	0.00 → High	0.09 → High	0.08 → High	0.20 → High	0.21 → High	0.00 → High	Commercial	Day	75.6
9	0.00 → High	0.06 → High	0.07 → High	0.00 → High	0.02 → High	0.00 → High	Commercial	Day	78.4
10	0.00 → High	0.22 → High	0.13 → High	0.20 → High	0.21 → High	0.00 → High	Commercial	Day	68.8
11	0.00 → High	0.01 → High	0.10 → High	0.20 → High	0.13 → High	0.12 → High	Commercial	Night	58.2
12	0.00 → High	0.00 → High	0.08 → High	0.00 → High	0.18 → High	0.00 → High	Commercial	Night	71.2
13	0.00 → High	0.20 → High	0.15 → High	0.00 → High	0.14 → High	0.16 → High	Silence	Day	56.1
14	0.00 → High	0.14 → High	0.11 → High	0.00 → High	0.14 → High	0.00 → High	Silence	Day	59.1
15	0.00 → High	0.23 → High	0.16 → High	0.00 → High	0.14 → High	0.16 → High	Silence	Day	54.8
16	0.00 → High	0.21 → High	0.15 → High	0.00 → High	0.14 → High	0.16 → High	Silence	Day	55.9
17	0.00 → High	0.01 → High	0.09 → High	0.40 → High	0.16 → High	0.05 → High	Silence	Night	48.5
18	0.00 → High	0.01 → High	0.09 → High	0.60 → Low	0.16 → High	0.05 → High	Silence	Night	44.4
19	1.00 → Low	0.79 → Low	0.90 → Low	0.40 → High	0.65 → Low	1.00 → Low	Industrial	Day	61.7
20	1.00 → Low	0.73 → Low	0.82 → Low	0.60 → Low	0.64 → Low	1.00 → Low	Industrial	Day	67.8
21	0.83 → Low	0.65 → Low	0.73 → Low	0.40 → High	0.64 → Low	1.00 → Low	Industrial	Day	71.3
22	0.00 → High	0.53 → Low	0.59 → Low	0.60 → Low	0.64 → Low	0.90 → Low	Industrial	Day	75.1
23	0.83 → Low	0.64 → Low	0.83 → Low	0.40 → High	0.61 → Low	0.48 → High	Industrial	Night	70.9
24	1.00 → Low	0.82 → Low	0.92 → Low	0.20 → High	0.49 → Low	0.88 → Low	Industrial	Night	53.8

FIGURE 17: Prediction of low noise level using different ML models for the input instances with probability values (0.001–1.00).
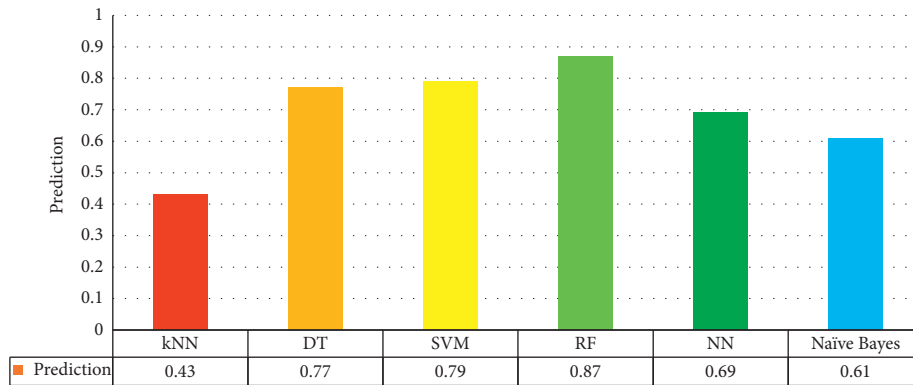

FIGURE 18: Graphical representation of prediction of low noise level using different ML models for the input instances with probability values (0.001–1.00).

prediction for an ML model. From Figures 15 and 16, it is observed that DT shows a higher average probability, 1.0, in predicting the *high noise level* than other models. RF shows the average probability of 0.94, NV shows an average probability of 0.68, k-NN shows average probability of 0.80, SVM shows an average probability of 0.87, and NN shows an average probability of 0.85. Hence, DT is better in predicting the high noise level. From Figures 17 and 18, it is observed that DT shows an average probability of 0.77 in predicting the *low noise level*, whereas RF shows an average probability of 0.87, NV shows an average probability of 0.61, k-NN shows an average probability of 0.43, SVM shows an average probability of 0.79, and NN shows an average probability of 0.69. Hence, RF is better in predicting the low noise level. From the results shown in Figures 15–18, it is concluded that the supervised ML models predict the noise levels (high/low) in a well manner with high probability of prediction.

4. Conclusion

The noise level in Bhubaneswar city during Dussehra 2020 has been studied using different supervised machine learning (ML) models, such as Decision Tree (DT), Neural Network (NN), k-Nearest Neighbor (k-NN), Naïve Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF), and has been simulated using Orange 3.26 data analytics environment. This empirical analysis demonstrated that DT and RF have a classification accuracy of 0.925, which is higher than that of other considered ML models. Aside from that, it is shown that the average probabilities of high and low noise levels being predicted in testing dataset are 1.00 and 0.87 for DT and RF, respectively. With the greatest machine learning models, we can anticipate noise levels in specific areas, such as residential, industrial, and commercial areas, in order to address the sustainability issue to the society for better life cycle. Furthermore, the key advantages of this study are the ability to classify and manipulate noises coming from many sources and to predict health issues in the near future. Due to the short dataset and the requirement for continual monitoring, this approach has significant drawbacks. Eventually, we plan to use this approach for estimating the noise levels in other regions by gathering more data by imposing recent deep learning (DL) based approaches, and different sensors in terms of IoT device can

be put at the location to continuously detect sound levels and save the data in a database like cloud. Furthermore, the detailed work has been planned to collaborate with government of Odisha (forecasting department) for future betterment in noise pollution monitoring in the real life environment.

Data Availability

Data are available on request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

References

[1] Smart city bhubaneswar, https://www.smartcitybhubaneswar.gov.in/know-bhubaneswar.

[2] Google Map, Bhubaneswar, https://www.google.com/maps/place/Bhubaneswar,+Odisha/@24.5960144,71.0428691,4.71z/data=!4m5!3m4!1s0x3a1909d2d5170aa5:0xfc580e2b68b33fa8!8m2!3d20.2960587!4d85.8245398.

[3] P. Cameron, D. Robertson, and J. Zaks, "Sound pollution, noise pollution, and health: community parameters," *Journal of Applied Psychology*, vol. 56, no. 1, pp. 67–74, 1972.

[4] A. K. Saha, S. Sircar, P. Chatterjee et al., "A raspberry Pi controlled cloud based air and sound pollution monitoring system with temperature and humidity sensing," in *Proceedings of the IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, Nevada, USA, 2018.

[5] "Environment monitoring data, state pollution control board, odisha (SPCBO)," 2021, http://ospcboard.org/environmental-monitoring-data/.

[6] Google Map, "Bhubaneswar locations," 2021, https://www.google.com/maps/@20.2858707,85.8273356,14.08z.

[7] S. K. Bhoi, S. K. Panda, K. K. Jena et al., "Prediction of diabetes in females of pima Indian heritage: a complete supervised learning approach," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, pp. 3074–3084, 2021.

[8] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: a review of classification techniques," *Emerging Artificial Intelligence Applications in Computer Engineering*, vol. 160, pp. 3–24, 2007.

[9] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 2016.

[10] N. Nowshin and M. S. Hasan, "Microcontroller based environmental pollution monitoring system though IoT implementation," in *Proceedings of the 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, Dhaka, Bangladesh, 2021.

[11] A. C. Pradhan, B. Kumar Swain, and S. Goswami, "Road traffic noise assessment and modeling of sambalpur city, india: a comprehensive, comparative and complete study," *Journal of Ecophysiology and Occupational Health*, vol. 12, no. 3/4, p. 51, 2012.

[12] B. K. Swain, S. Goswami, and M. Das, "A preliminary study on assessment of noise levels in indian offices: a case study," *Asian Journal of Water Environment and Pollution*, vol. 11, no. 4, pp. 39–44, 2014.

[13] B. K. Swain, S. Goswami, and M. Das, "Assessment of highway noise and predictive models along NH-316," *Asian Journal of Water Environment and Pollution*, vol. 13, no. 2, pp. 97–105, 2016.

[14] B. K. Swain, S. Goswami, and S. K. Panda, "Road traffic noise assessment and modeling in bhubaneswar, india: a comparative and comprehensive monitoring study," *International Journal of Earth Sciences and Engineering*, vol. 5, no. 1, pp. 1358–1370, 2012.

[15] B. Swain and S. Goswami, "Soundscapes of urban parks in and around bhubaneswar and puri, odisha, india: a comparative study," *Pollution*, vol. 4, no. 1, pp. 93–101, 2018.

[16] S. Goswami, B. K. Swain, and S. K. Panda, "Assessment, analysis and appraisal of road traffic noise pollution in rourkela city, india," *Journal of Environmental Biology*, vol. 34, no. 5, pp. 891–895, 2013.

[17] A. K. Sahu, S. K. Nayak, C. R. Mohanty, and P. K. Pradhan, "Traffic noise and its impact on wellness of the residents in sambalpur city–a critical analysis," *Archives of Acoustics*, vol. 46, pp. 353–363, 2021.

[18] A. K. Sahu, S. K. Nayak, J. K. Mohanty, P. K. Pradhan, and C. R. Mohanty, *Advances in Mechanical Processing and Design*, Springer, Berlin, Germany, 2021.

[19] A. K. Sahu, M. Pradhan, C. R. Mohanty, and P. K. Pradhan, "Assessment of traffic noise pollution in Burla town, India; an inclusive annoyance study," *Sound and Vibration*, vol. 54, no. 1, pp. 27–42, 2020.

[20] A. K. Sahu, P. K. Pradhan, S. K. Nayak, M. Pradhan, and C. R. Mohanty, "Evaluation and modeling of traffic noise in berhampur city, India," *Fluctuation and Noise Letters*, vol. 19, no. 4, Article ID 2050044, 2020.

[21] Orange data analytics tool, https://orangedatamining.com/.