IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Predictive Framework for the Urban Environment Monitoring using Artificial Intelligence and Wireless Sensor Network

Rashmi S Bhaskar
Research Scholar, Visvesvaraya Technological University, Karnataka, India

Dr. Veena S Chakravarthi
Professor, Department of Electronics and Communication Engineering, BNMIT, Bengaluru, Karnataka, India

***Abstract*—The advanced systems based on the wireless sensor network, artificial intelligence and the eco-system of the communication protocols provide possibilities of building more efficient, robust and flexible system for the urban environment monitoring. The existing and obstinate constraints of the communication channel have various limitations of the real-time monitoring based on the network performance of the data delivery through the wireless sensor network deployments in the urban environment. These challenges can be handled by the use of smart and intelligent systems using predictive modeler by incorporating the machine learning aspects of the artificial intelligence by means of behavioral study of the historical data of the environment conditions in the urban context. Another critical challenge is that the deployment strategy for specific application demands a customized architecture of the wireless sensor network setup. To alleviate this challenge, this paper basically provides a generalized architecture design approach for the urban environment monitoring with the use-case study on the air quality dataset, water quality dataset, urban power usage dataset and storm risk prediction dataset so that the innovative solution approach becomes adaptive to the varied application context.**

***Keywords*—Wireless Sensor Network, Urban environment, Artificial Intelligence, Machine Learning, Predictive Model**

I. INTRODUCTION

The sensor nodes of the wireless sensor network basically transform the real-time environmental changes into data. The customization of the sensor nodes largely depends upon the on-situ conditions of the deployment whether it is under earth, over-earth or inside the water or chemicals and different exposures to it [1]. The eco-system of the urban region is different than the rural as the in urban region there exist factories, airport, transport system, educational institution, systematic water supply & sewerage, hospitals, government offices, road traffic, shopping mall etc. to cater different facilities and services to the large population in the defined area [2-3]. The placement of the sensor network, selection of the communication protocols, interferences due to obstacles, channel conditions and another deployment and network related issues poses a server challenge to meet the specific goals of monitoring in a cost effective and accurate way [4]. Figure - 1 illustrates the three different views as a evolution to the deployment architecture of the wireless sensor network, where the latest architecture provides a tremendous opportunity to deal with the data and make the urban environment monitoring system (UEMS).

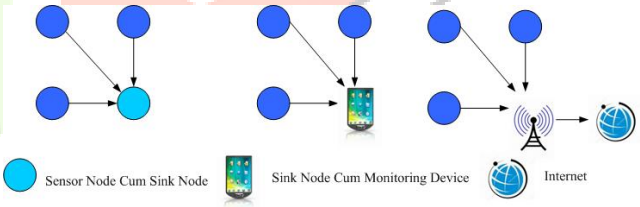

Figure 1: Evolution of wireless sensor networks deployment architecture

The different architecture of deployment is used as per the need of the application and the deployment conditions, however synchronization of the wireless sensor network with the internet provides better possibilities of data analysis using predictive molders as it can connect to cloud and end user devices and distributed monitoring systems [5]. An efficient learning based predictive models are desirable in the urban environment as the hazard of interferences are higher due to the distortions caused in the communication in the context of the urban environment as compare to the rural environment [6]. The increasing population of the urban area brining more and more congestions as well as cause of noises and pollution, thus the solution schemes shall focus on the cost effective and efficient systems which is the core goal of the researchers focus. And as a result, many use cases and unconventional applications are emerging out that caters services and solutions to the different walks of the life for urban population. Few such popular applications include water-supply and sewage management, structural health monitoring systems, disaster and accident management, law enforcement and criminal activity tracking along with various environmental monitoring system [7-8]. These

solutions design considers the optimality with the constraints including the noise and the space. The figure-2 provides a snap view of the application scenarios in the urban context. There are various work being carried out on specific applications for the urban environment monitoring on the dataset obtained from the sensor network using machine learning as described in the section -2 of review of literature. Based on the review analysis it is realized that a generalized framework is require to deals with the different application aspect for different dataset.

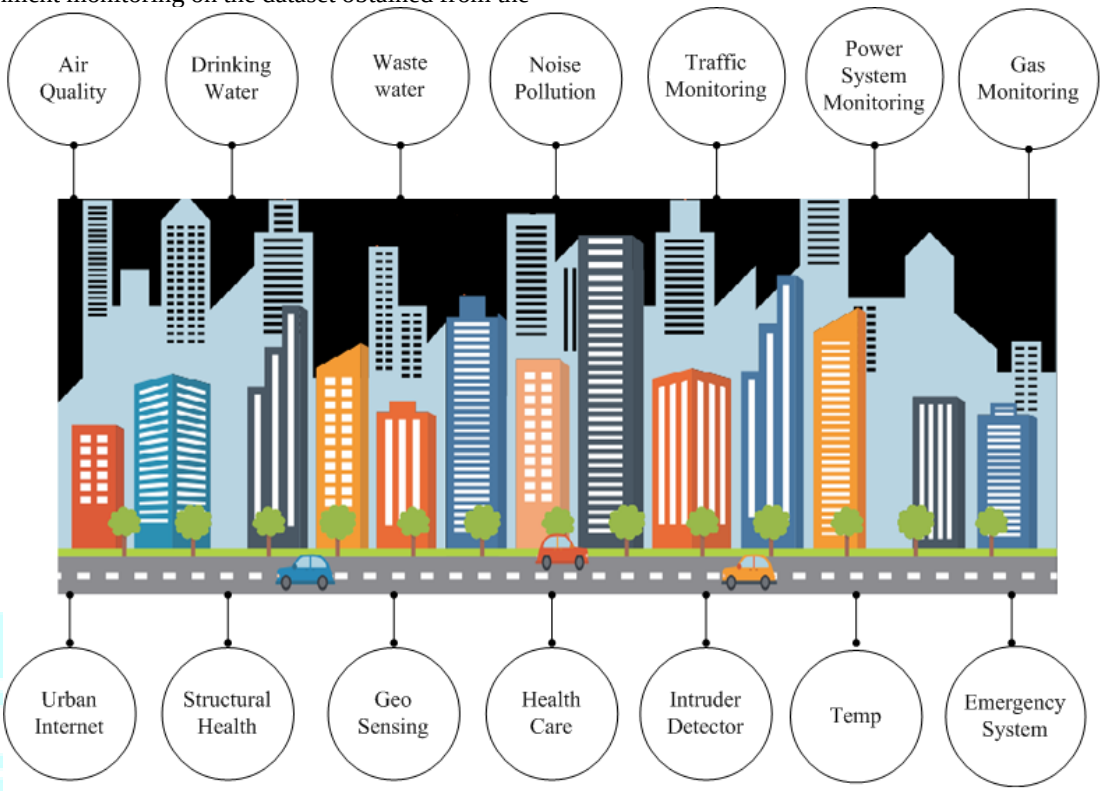

Figure 2: Application scenarios in the urban context

Therefore, this paper presents a predictive framework for analysing adequate placement of the sensor nodes for capturing vital environmental events to benefit planning and decision making processes that contributes to more habitable and sustainable environemnts. The development of the predictive framework is based on four classification and regression analysis techniques. The regression analysis is implemented for predicting continuous dependent variables and classification is performed for discreate dependent data. To validate the proposed framework the study considers four different dataset that includes: i) water quality, ii) Storm Risk, iii) Air Quality and iv) urban power consumption that undergoes through different classification and regression analysis. The major contribution of the proposed work are multiple described as follows:

- An effective framework is designed as a support system for the urban environment monitoring.
- The sensory data quality is predicted with respect to the precise placement of senors to collect high quality data.
- Data visualization and preliminary analysis is carried out to understand the data charaterstics towards effective data modeling as an input to the predictive model.
- Implementaion of suitable classification and regression techniques for processing continous and discreate data related to the different environmental events.

Based on the output statistics a suitable classication and regression technique is selected according to which most suitable type of deep learning mechnaims can be further designed and implemented towards a reliable and automated system which can efficiently capture and analyze data on the environment in order to avoid any potential risks.

The remaining sections of this paper are organized as follows: Section II presents related work on specific applications for the urban environment monitoring; Section III presents schematic architure of the proposed predictive framework; Section III presents visualization and preliminary analysis of the dataset; Section IV presents a brief details on the the classification and regression technique adopted in the proposed framework. Section V presents the outcome obtained and performance analysis of the proposed system; and finnaly Section VI concludes the entire work.

II. Related Work

For the human health it is quite essential to monitor air quality as in the urban conditions the changes in environment takes place due to the air pollution caused by the industrialization and traffic pollution. The pollutant like PM , excessive of SO2, CO2,O3 and NO2 require actions on the real-time basis. In the work of (Amuthadevi et al, 2021) machine learning methods like LSTM is used for the real-time corrections on time-series data[9]. Machine learning approach is being used for the drinking water treatment [10] the survey work by (Li L et al, 2021) describes in detail about the various methods of AI used and reveals the fact that how exactly the AI based methods helps to categorizes the containment and another decision systems for the plants of drinking water. Additionally, towards the waste water management using machine learning and AI approach many works can be found in the work of (Heo S et al

, 2021) [11] , Biological waste water using AI in the work of (Sundui et al, 2021)[12-15].

One of the important urban pollutions is noise pollution that severely affect both physical and mental health of the people. It is being found that a noise map generated with the heap of the flow of the traffic and the techniques that evaluates the propagation of the noise can be correlated to predict or identify the affected population and in accordance remedial actions can be taken. However, there exist certain limitations to identify the exact and accurate impact zones as the categorization of different types of noises are very hard through these methods, thus in the work of (Alvareset al, 2021) the accuracy is improvised by considering the walking surveys and lower-level dataset from the sensor network and used Machine learning for the categorization [16]. Many another work is found in literature that focuses on the noise pollution in different context and different aspects which machine learning is broadly used on the sensor-based data. Few of such recent work includes by (Zhao et al, 2021) [17] , (Chandra et al, 2021) [18] and (Van et al , 2021)[19].

In the urban context, setting up the intelligent transport system is the primary requirement for both safety and comfort. The collaborative research of VANET and machine learning paves the foundation to meet this goal as described in the work of (Khatri et al)[20], (Khan et al, 2021)[21], (Li C et al, 2021)[22]. Power system is the core backbone of the urban infrastructure management and it is going through a paradigm shift. Many health related system such as mosquito control is designed using machine learning by (Joshi et al, 2021)[23] , however another works towards the power system management using machine learning includes works by (Yang et al, 2021) for protection and control [24] , (Alimi et al) for security and stability [25] , (Malbasa et al, 2017) for voltage stability [26], (Karimipour et al, 2019) for cyber-attack detection in smart grid[27] and (Tian et al , 2021) for energy categorization of buildings in urban context[28].

Fuel economy and gap monitoring is another requirement to be monitored and optimized for the benefit of the urban population. The driving behaviour impact the fuel economy, in the work of the (Kim et al , 2021) , machine learning is used on the data from the drivers driving behaviour [29]. The gas monitoring using machine learning is being extensively studied in the work of [30] and [31]. The internet is the life-line after electricity for the urban development, however there always exist a threat to it by the extremist and un-social elements. In the work of (Mashechkin et al, 2019), machine learning is used to classify the user type on the internet in the urban context based on the pattern identification [32]. One another important aspect in the urban context is to continuous monitoring of the structures such as multi-storeyed building, bridges and another important establishment. The extensive survey by (Yuan et al, 2020), clearly describes that machine learning plays a vital role to build an application for structural health monitoring from the sensor data [33]. Land use monitoring is quite helpful application for the urban administration and in direction the work by (Kafy et al, 2021) proposes a prediction model using machine leaning using geo sensing data that helps to comply the requirement of the sustainable urban development [34].

Wireless Sensor network plays an important role for the health care monitoring in urban environment towards both proactive and reactive approach towards healthcare management. In the work of (Ogunyemi et al, 2021) deep learning is used for identifying the diabetic retinopathy [35]. For the benefit of the urban population those suffering from cardiac disorder, the authors (Alghamdi et al, 2020) proposes a model based on artificial intelligence for the smart city health care management especially using deep learning approach [36]. Similarly, machine learning is broadly used for intrusion detection [37-39] , temperature monitoring[40-42] and emergency system[43-45].

III. Problem Description

An environmental monitoring system refers to a computational model which captures the vital events of the different environmental factors such as air quality, presense of carbon content in the air, smoke, humidity, quality of water temperature, humidity, city power consumption, storm prediction, dust particle and many more. It also provides a essential statistical analysis towards effective planning decision making process about critical situations.

Based on the review of the existing research work, it has been analysed that literature is rich containing variety of the schemes and solution concerning environmental monitoring. But no research study concentartes on assessing sensor placement concerning capturing of reliable data in case of water monitoring and storm risk prediction. As it is obivious that unprecise placement of sensor nodes can mislead environmental planning and critical deciosn making processes. For example, if an water sensor nodes placed incorrectly then it captures false data and the results may be misleading, the insight produced are errorsome and whole environmental system will be affected. Therefore, to avoid such scenario, a precise placement strategy need to be adopted which prevent capturing of ambiguous information and ensures data integrity. It has also been studied that no existing works have suggested a unified framework that can support processing of varied dataset having massive amount of environmental events produced by the sensor nodes. Also, the implementation of existing model for environment monitoring is carried out with complex mechanism that involes recursive operation prone to the huge computational complexity. Thus hinders their scope of applicability in real-time scenario. Another important factors analysed from the literatures is that, the existing works have not shown any evidence or fact that on what basis they have implemented or developed their predictive model. Since, there are variety of learning models avaibale such as machine learning (ML) and deep learning (DL) technique and each have their own advantages and limitations. Therefore, selection of suitable machine learining or deep learning technique becomes challenging taks. In order to build an efficient and reliable predictive model selection of ML or DL should be done on the empirical and evidential analysis. These are the few facts that motivates us to suggest a preditive framework as a support system for designing a reliable and adaptive environment monitoring system.

IV. Propsoed Framework

The proposed study aims to provide an effective unified framework that acts a support system towards designing and developing the adaptive and relible urban environment monitoring system contributing more livable, healthy and sustainable surroundings. The schematic architecture of the proposed framework is shown in Figure 3.

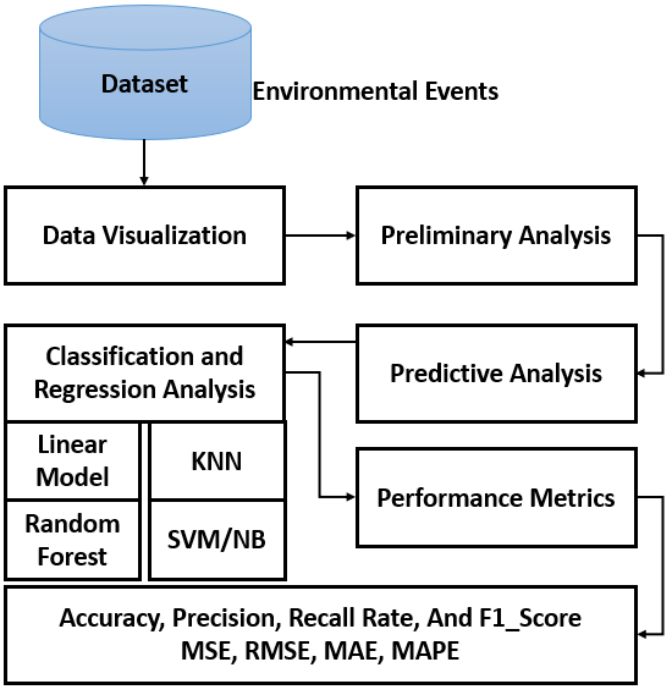

Figure 3: Schematic architecture of the proposed framework

The proposed framework considers a multiple datasets related to different environmental factors such as air quality, water quality, power consumption and storm prediction. The initial operation carried out in the proposed system is the visualization and preliminary analysis of the dataset, which gives insignt and provides a better understanding about the dynamics and nature of the dataset (continuous and discreate). Also, this operation helps to determine the predictors (input) and response variable (output) towards modelling the predictive model. The study utilizes different types of machine learning techniques to carry put predictive analysis based on the requirement. The different machine learning based classification technique are applied for discreate dataset and in the similar way different regression analysis is applied for the continuous dataset. Based on the outcome statistics and performance measure s suitable technique can be selected towards opting or deciding the designing of advance learning modle based aspative and reliable environment monitoring system.

4.1 Water Quality Dataset

This section presents dataset visualization and perform exploratory analysis towards building an understanding about the water quality dataset [46].

A) Dataset Visualization

In this dataset samples are collected for water quality from three sources that includes i) Streams, ii) Lakes and iii) Puget Sound. Table 1 illustrates the name of the identifier of the data, and the datatype.

Table 1: Name of the identifier of the data, and the datatype

Sl. No	Data identifier	Datatype
1	sample_id	Integer
2	grab_id	Integer
3	profile_id	Integer
4	sample_number	String
5	collect_datetime	datetime
6	depth_m	Decimal
7	site_type	string
8	area	string
9	locator	string
10	site	string
11	parameter	string
12	value	decimal
13	units	string
14	**Quality_id (output)**	integer
15	lab_qualifier	string
16	mdl	decimal
17	rdl	decimal
18	text_value	string
19	sample_info	string
20	steward_note	string
21	replicates	integer
22	replicate_of	integer
23	method	string
24	date_analyzed	date
25	data_source	string

B) Preliminary Analysis

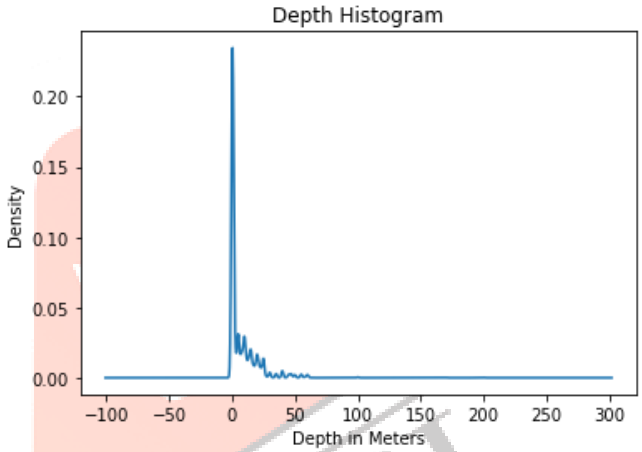

Figure 4: Histogram plot for water depth analysis

The above Figure 4 demonstartes the probability distribution of water depth which shows most of the water is in shallow depth. There are some places where water is 50 (164 feet) meters deep. The next Figure 5 presents an analysis of water samples taken from the different regions of the city.

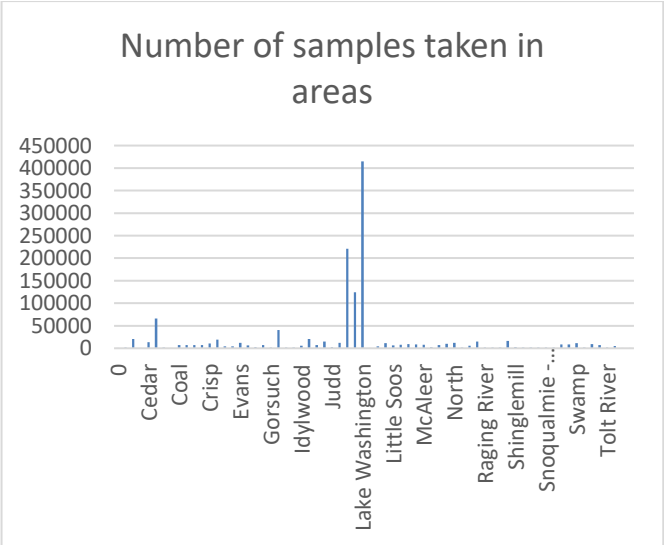

Figure 5: Count of water samples taken from different rivers

The above graph shows number of samples taken in various areas of the city. More samples are taken from lakes and hence we may expect good data from those sensors however there are also more number of shallow waters and hence due to that the data collected in these regions might be of lower quality and this might cause some imbalance in the data.

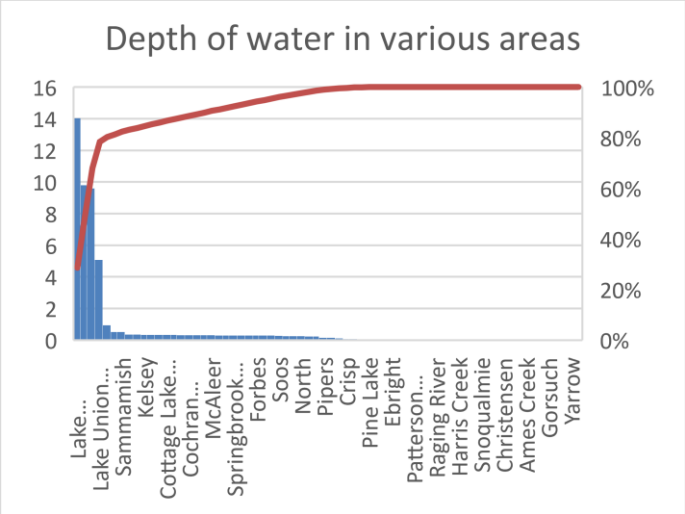

Figure 6: Depth of water in different rivers or lakes

From Figure 6 it can be seen that, shallow waters are more in case of creeks and other areas where as deep waters can be seen in lake Washington and lake Sammamish. However as it can be observed from the orange line the number of samples are in less percentage from these places where as the most number of data is collected from other places where there are shallow waters. The quality id for the collected data is mentioned below:

0 – Quality Unknown
1 – Good Data, Passes Data Manager QC
2 – Provisional Data, Limited QC
3 – Questionable/Suspect
4 – Poor/Bad Data
5 – Value Changed (see Steward Note)
6 – Estimated Value
9 – Missing Value

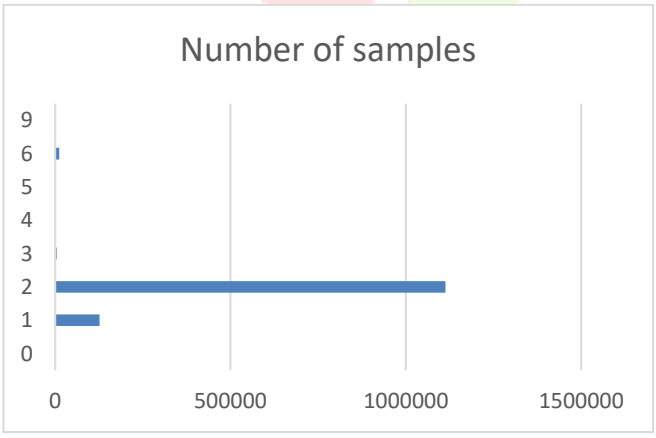

Figure 7: Analysis of count of water samples

From Figure 7 it can be observed in above graph, even though, there is a huge number of provisional data. However it is acceptable by QC according to the documentation. Provisional data only means that final count may differ. As it can be observed that high quality data is also available and poor quality data is very limited in number.

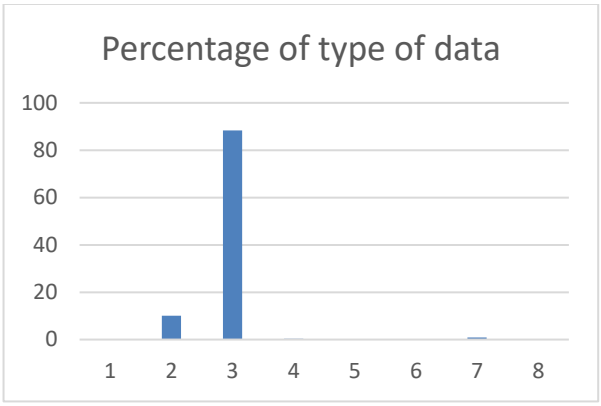

Figure 8: Analysis of water data quality type

The graph trend shows percentage of acceptable data for each category as mentioned in Figure 6. The quality of third data sample exhibits higher percentage of being poor quality or misleading data. This shows a clear imbalance in the data. However, in this framework we are using any type of regulizer and we want to see the performance of the algorithms on raw data to decide on the further development strategy for ML development.

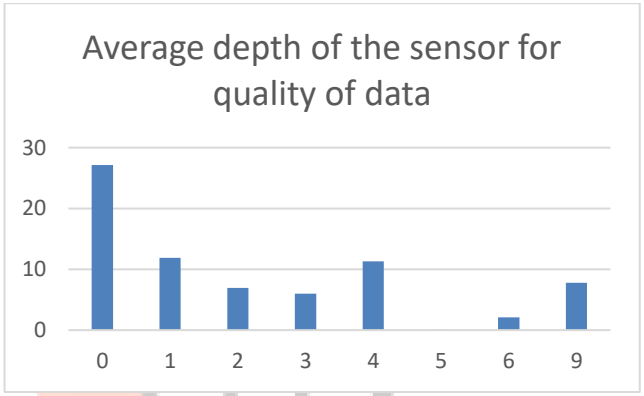

Figure 9: Analysis of average depth of the sensor

The above Figure 9 shows average depth of sensors with respect to quality of captured data.

4.2 Storm Risk Dataset:

This section presents dataset visualization and perform exploratory analysis towards getting insight about the storm risk prediction [47].

A) Dataset Visualization

In this dataset, the data is collected that represents regions where the risk of storm tides exists due to hurricanes. Table 2 illustrates the descriptions of the identifier of the data, its description and the datatype.

Table 2 : Name of the identifier of the data and the datatype

Sl. No	Data identifier	Datatype
1	OBJECTID	Integer
2	JOIN_COUNT	Integer
3	TARGET_FID	Integer
4	**HES_ZONE (output)**	Integer
5	CONTOURLEN	Integer
6	SHAPEAREA	Integer
7	SHAPELEN	Integer
8	geometry	String
9	coordinates	Integer

B) Preliminary Analysis

Figure 10: Geoplot for highly storm prone area

The analysis from the above graph exhibits that Washington area is being higly considered for storm risk analysis.

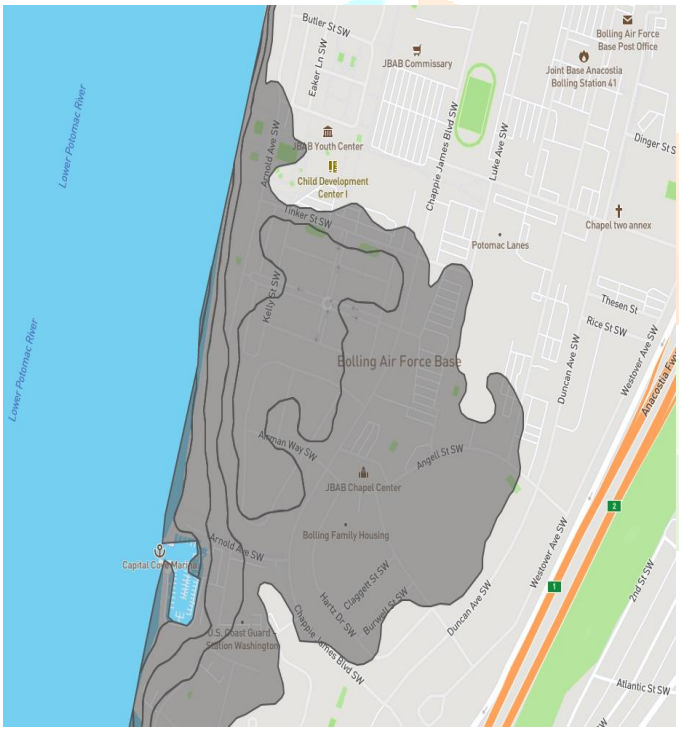

Figure 11: First high risk area in Washington near Capital Cove harbor

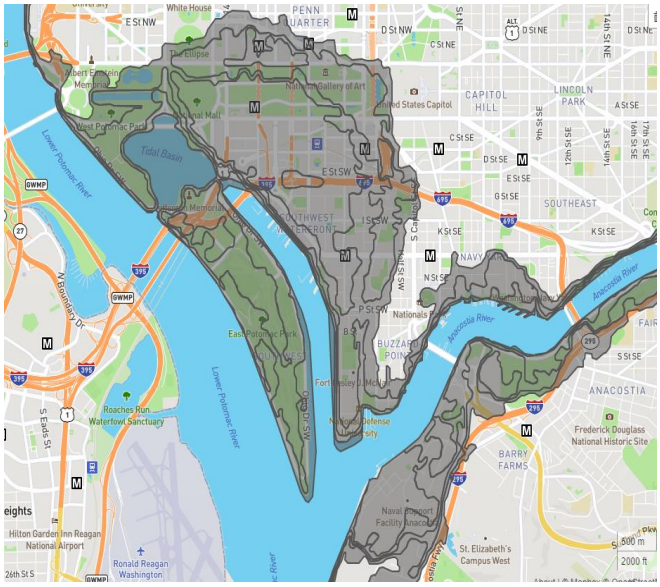

Figure 12: Second high risk region in Washington city near joining of Anacostia river and Lower Potomac river

This particular region has been hit by the storm several times. This is the place where most of loss of life and property has happened. This is due to the fact that this region is closer to ocean and this is the pace river The next figure demonstrates high risk area and low risk area in the Washington.

Figure 13: Low storm risk regions in Washington near Fredrick Douglas Memorial bridge

4.3 Air Quality Dataset

This section presents dataset visualization and perform exploratory analysis towards getting insight about the air quality analysis [48].

A) Dataset Visualization

In this dataset, it provides the ambient air quality as per the standard from different air monitoring points for SO2 , NO2, CO and PM2.5 & PM10. Table 3 illustrates the descriptions of the identifier of the data, its description and the datatype.

Table 3 : Name of the identifier of the data, and the datatype

Sl. No	Data identifier	Datatype
1	date	string
2	time	decimal
3	**co_mg_m3 (ouput)**	float
4	8_hr_roling_avg	Float
5	flag	String
6	comment	String
7	co_mg_m3_2	Float
8	8_hr_roling_avg_2	Float
9	flag_2	string
10	comment_2	string

B) Preliminary Analysis

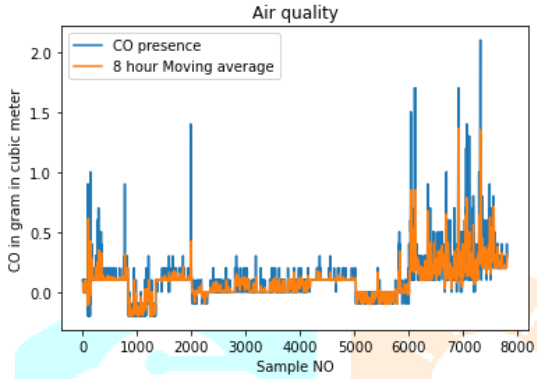

Figure 14: Presence of carbon monoxide in every 8 hours

From the above Figure 14, 8 hour moving average of pollution level in Dublin city can be seen.

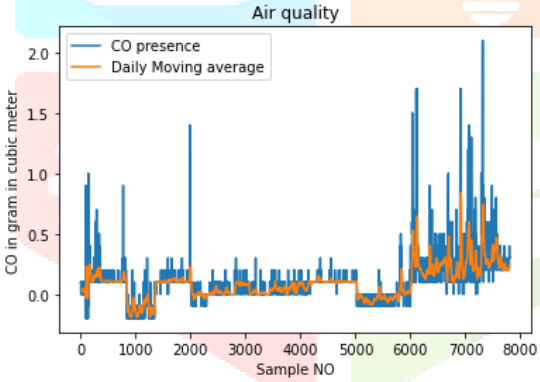

Figure 15: Presence of carbon monoxide daily moving hours

In above Figure 15, daily moving average can be seen. Both the moving averages show that the day time pollution is getting higher during the end of 2012.

4.4 Power consumption

This section presents dataset visualization and perform exploratory analysis towards getting insight about the amount of electricity consumption in the city [49].

A) Dataset Visualization

In this dataset, it provides power consumption of the city of Seattle. It has totally got 45 column headers few of the column headers are as in Table 4.

Table 4 : Name of the identifier of the data, and the datatype

Sl. No	Data identifier	Datatype
1	PrimaryPropertyType	String
2	PropertyName	String
3	TaxParcelIdentificationNumber	Integer
4	Location	String
5	CouncilDistrictCode	Integer
6	YearBuilt	Integer
7	Neighborhood	String
8	….	….
9	…	…
10	ENERGYSTARScore	Integer

B) Preliminary Analysis

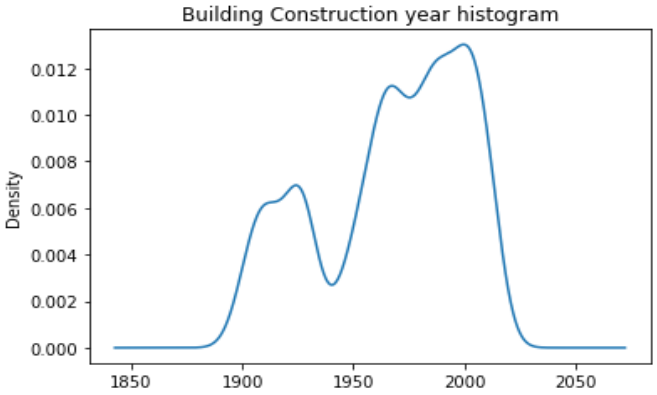

Figure 16: Histogram plot of building age

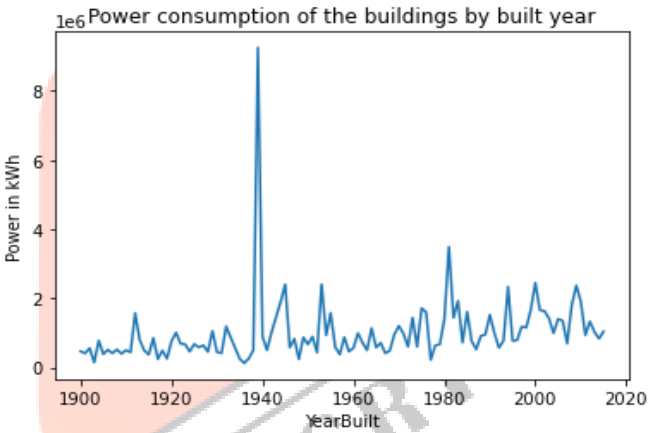

Figure 17: Histogram plot for closer analysis of building age

It can be seen that the average power consumption of modern buildings are higher. The spike in the center can be ignored as that was the time of world war and some major factories were built during that time.

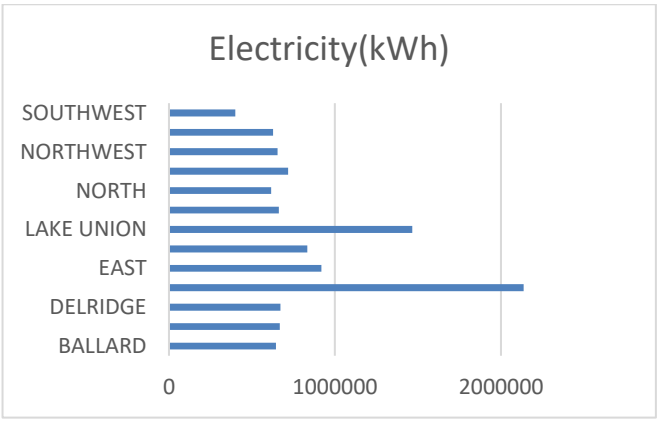

Figure 18: Histogram plot for closer analysis of building age

Above Figure 18 shows average power consumption of various areas. The above graph shows that the average power consumption in downton area is higer compared to rest of the areas. Hence it can be cunclded that the area in which the

building is located is an important factor in deciding the power consumption of the building. This is due to the fact that the building located in the downtown area run for more time and they are mainly shops and malls.

4.5 Predictive Techniques

Linear regression: Linear regression is the simplest of all regression algorithms. This simply multiplies all the predictors with the weights and a single bias is considered. The Linear regression assumes that the relationship between the predictors and response are linear in relationship. However, the relationship between predictors and response is never linear in case of real-world data. However, for any data which is close to being linear, this model performs really well. Better performance of this algorithm shows that complex statistical models are not needed in order to perform predictive analysis on the data.

SVR: Support vector regression or SVR is a type of ML algorithm which uses probabilistic models in order to find the value of the response variable. The support vectors are the vectors in the n dimension space of the predictor variables. These vectors prevent the errors from happening. The value of the response variable has a higher probability of being present between the two support vectors. These support vectors represent the probability of the value. This is a probabilistic model which always predicts the value with the help of a probability model. Unless specified, the SVR uses normal distribution to find the suitable value for the predictor.

Random forest regression: Random forest is nothing but a group of decision trees and the value of the response variable is always predicted by the leaf node of the several decision trees present in the random forest model. The final output is the average of all the values predicted by each of the decision trees present in the random forest model. This model performs the best when the input is mostly consisting of discrete values.

K neighbor regression: This is a special kind of k neighbor model where each node in the N dimensional space consists of a value instead of a category. The value of the present node is interpreted as the average value of the K nearest neighbor nodes to the present node. This model is best suited for prediction of special data such as geological sensor placement data.

Logistic Regression: This is the basic classification algorithm which is close match to the linear regression since it uses the generalized linear model and uses sigmoid as the connection function. This is the basic classification ML which use 1 weight for each variable and a bias. This model works best when the relationship between predictor and response data is linear.

Gaussian NB: This is the probabilistic classification model which uses bayes theorem to predict the outcome. The name Gaussian suggests that it uses Gaussian distribution which is best suited for probability based classification problems.

Random forest classification: This is very similar to random forest regression however the leaf node contains a class instead of a value. The results of the several decision trees are aggregated by considering the count of classes instead of averaging.

KNN classifier: This is the well known classifier which considers the class of the N nearest nodes and the most repeated count of the neighbor nodes is considered.

V. Results and Performance Analysis

The predictive framework proposed acts as effective support system for building an efficient and automated model of the urban environmental monitoring system. The design and development of the proposed environemtal support system is carried out on the Anaconda tool using Python programming language. This section presents the outcome obtained and a comparative analysis in to show the effectiveness of the each implemented machine learning classification and regression techniques.

5.1 Performance analysis on Water quality dataset

In this section the outcome is shown for the classification techniques for the prediction of discreate dependent values for evaluating the reliability of the sensory data for monitoring and assessing the water quality. Table 1 shows quantified outcome obtained for the classification techniques in this context.

Table 5: Quantified Outcome

Algorithms	Accuracy	Precision	Recall	F1 score
KNN	0.83134	0.81252	0.8561	0.83374
Logistic Regression	0.83562	0.8524	0.82436	0.83814
Gaussian NB Classifier	0.91532	0.9142	0.91235	0.91327
Random Forest Classifier	0.9351	0.93143	0.9434	0.93737

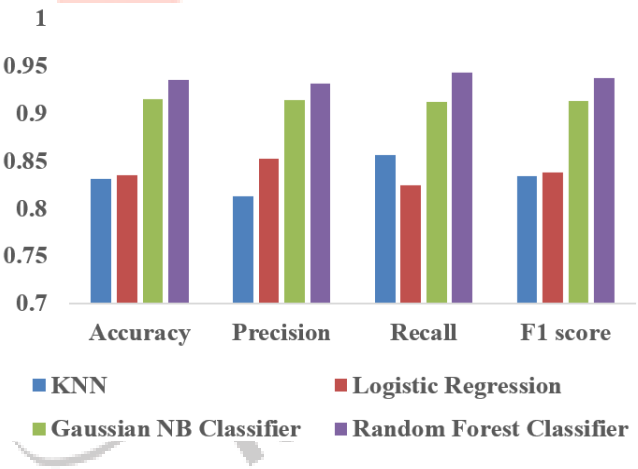

Figure 19: Comparative analysis

Above graph shows performance metrics of various algorithms for this dataset. Here random forest performs better than any other algorithm since the dataset contains more discrete values. As discussed earlier, when there are more discrete values are present in the input, better the performance of the decision tree based algorithms.

5.2 Performance analysis on Strom risk Dataset

In this section the quantified outcome and performance analysis is shown for the classification techniques for the prediction of discreate dependent values for evaluating the reliability of the sensory data for monitoring and prediction of strom risk. Table 6 shows quantified outcome obtained for the classification techniques.

Table 6: Quantified Outcome

Algorithms	Accuracy	Precision	Recall	F1 score
KNN	0.95243	0.98342	0.92412	0.95284
Logistic Regression	0.8185	0.815434	0.82353	0.81946
Gaussian NB Classifier	0.91012	0.91434	0.91456	0.91445
Random Forest Classifier	0.85232	0.82423	0.88234	0.85229

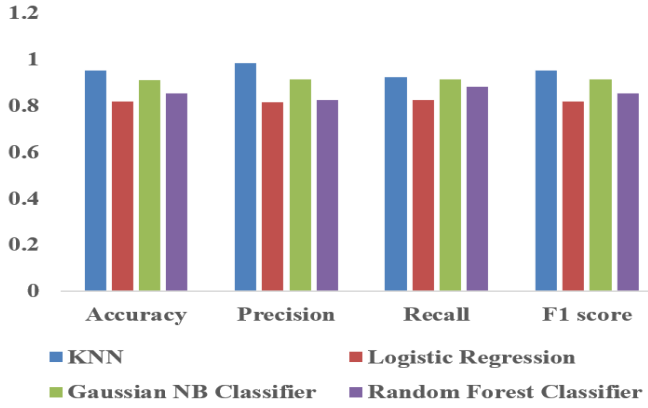

Figure 20: Comparative analysis

The analysis shown in Figure 20 are because of the fact that the data is geological and KNN is most suited algorithm for such purposes. The main data input here is the coordinates of the places. There are more risky places near harbors and airports. The more expensive properties are located closer to such risky places. This type of data is best interpreted in an N dimensional space. Hence KNN is the best suited algorithm here.

5.3 Performance Analysis on Urban Power Consumption

In this section the quantified outcome and performance analysis is shown for the classification techniques for the prediction of discreate dependent values for evaluating the reliability of the sensory data for monitoring and prediction of urban power consumption. Table 7 shows quantified outcome obtained for the classification techniques.

Table 7: Quantified Outcome

Algorithms	MSE	RMSE	MAE	MAPE
Linear Regression	404983	2012.42	1996.42	1.961332
K-neighbor Regression	208377	1443.53	1425.53	1.400475
Support vector Regression	104738	1023.42	1008.42	0.990696
Random Forest Regression	130537	1142.53	1131.53	1.111643

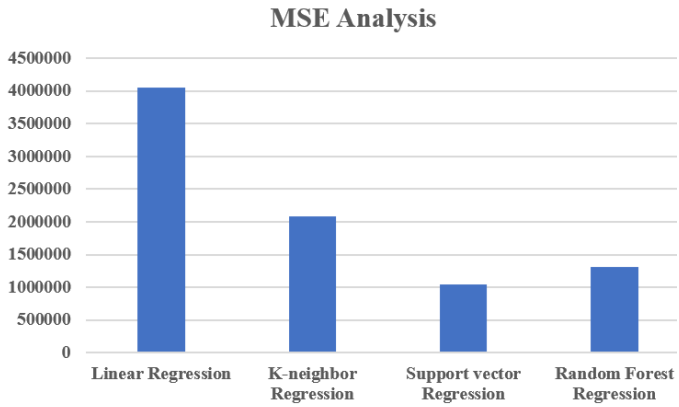

Figure 21: Comparative analysis in terms of MSE

In Figure 21, the analysis of mean square error is shown for data concerning power consumption. The MSE is a very sensitive error and less forgiving towards outliers. The power consumption of the buildings works best with the probabilistic models. The error is expected the least when the model is performing best.

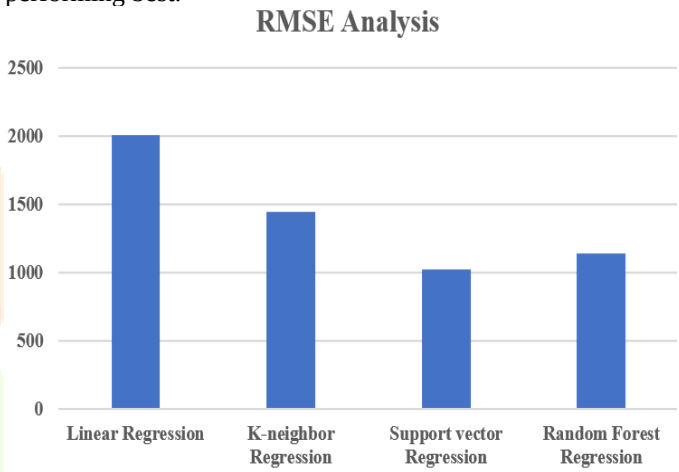

Figure 22: Comparative analysis in terms of RMSE

In Figure 22 RMSE of the algorithms confirm what MSE has shown since they are just the square root of the MSE, they are also sensitive to outliers.

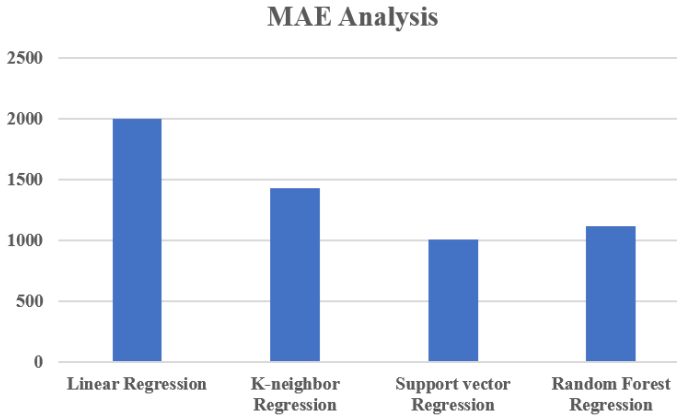

Figure 23: Comparative analysis in terms of MAE

In Figure 23 the performance in terms of MAE shows that there is not much outliers in the data. Since there is not much difference between RMSE and MAE, the count of outliers must be low in the data. However here also it is confirmed that the SVR is the best performing algorithm and hence the data is suitable to be processed with a probabilistic model.

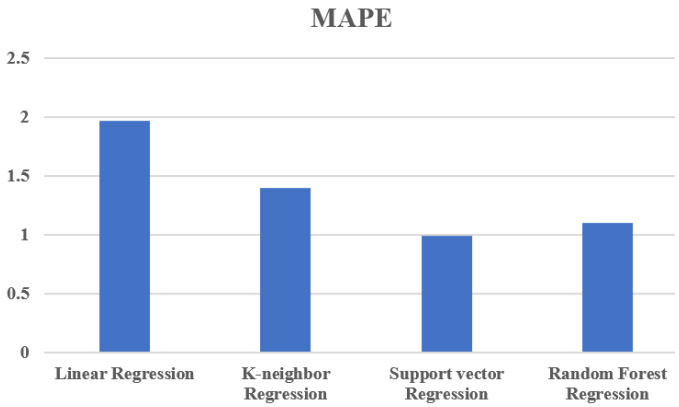

Figure 24: Comparative analysis in terms of MAPE

The analysis from Figure 24 shows Mean absolute percentage error of the algorithm confirms what is seen in case of MAE. The MAPE is mean absolute percentage error. Hence it shows the error in percentage. The error of SVR is around 1% which means that the user can expect a 1% deviation on the end result of this algorithm.

5.4 Performance Analysis on Dublin Air Quality Dataset

In this section the quantified outcome and performance analysis is shown for the classification techniques for the prediction of discreate dependent values for evaluating the reliability of the sensory data for monitoring and prediction of air quality (grams of carbon monoxide in cubic meter of air). Table 2 shows quantified outcome obtained for the classification techniques.

Table 8 Quantified Outcome

Algorithms	MSE	RMSE	MAE	MAPE
Linear Regression	7.37E-05	0.008587	0.00843	0.42154
K-neighbor Regression	3.79E-05	0.00616	0.00559	0.27960
Support vector Regression	1.91E-05	0.004367	0.00383	0.19173
Random Forest Regression	2.38E-05	0.004875	0.00253	0.12665

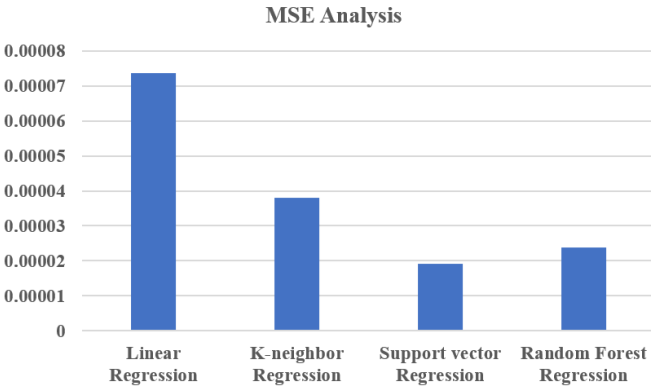

Figure 25: Comparative analysis in terms of MAPE

The air quality dataset is a time series dataset and also it depends on the probabilities of the previous data. In a time series, the data is considered as the probabilistic data. Same trend can be seen in results as well.

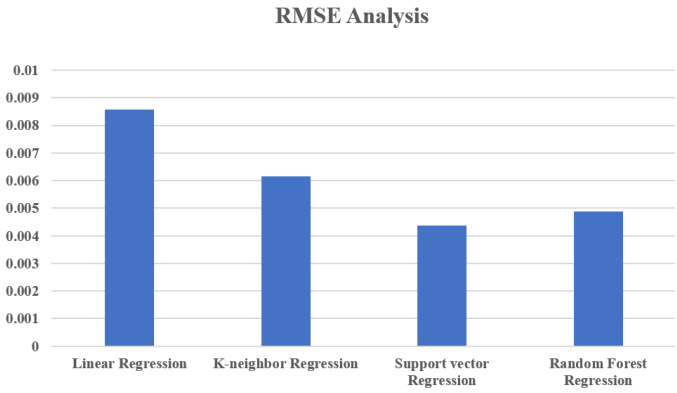

Figure 26: Comparative analysis in terms of RMSE

Since RMSE is square root of MSE, the same trend can be seen in the RMSE as well. With RMSE of 0.0045 in SVR, it shows that the probabilistic models are suitable here.

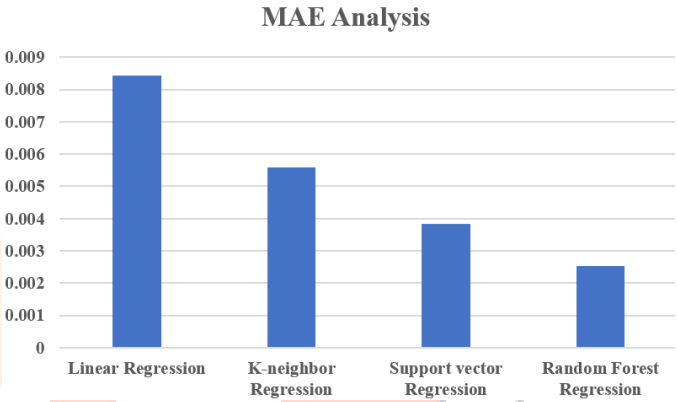

Figure 27: Comparative analysis in terms of MAE

As it can be observed, MAE in Random forest is less than RMSE and is changing. This is a clear indication of presence of outliers. Outliers can be seen in the basic analysis as at the end of 2012, there is high variation in air quality trend.

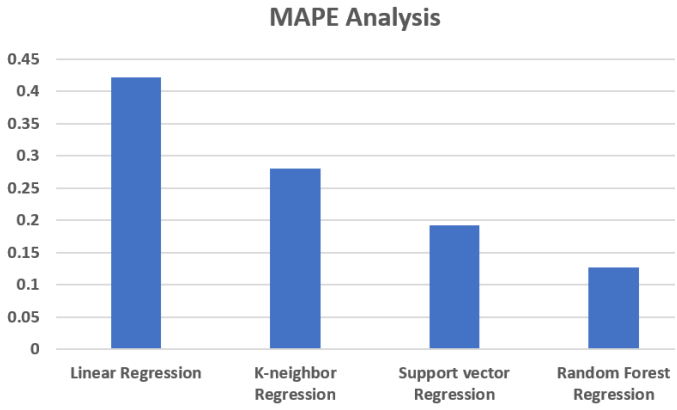

Figure 28: Comparative analysis in terms of MAPE

MAPE follows the same trend as the MAE and confirms the presence of outliers in the data.

VI. CONCLUSION

The proposed work suggested a predictive framework that contributes vital role in urban environment monitoring. The proposed framework integrates different machine learning models for executing classification and regressing analysis over the data obtained from sensor nodes deployed for monitoring different environmental factors such presence of carbon monoxide in the air, water quality, risk of storm in city, and usage of power in the city. A machine learning based regression analysis is carried out for the analyse quality of air and prediction of the power usage. On the other hand classification is carried out for prediction sensory data quality concerning water quality monitoring and storm risk prediction. The design of the framework is flexible to process multiple dataset of different environment data it is not limited to single or specific dataset. In the future work, based on the performance analysis and output statistic a suitable deep learning technique will be designed for reliable and efficient environment monitoring system.

REFERENCES

[1] Khan MW, Salman N, Kemp AH, Mihaylova L. Localisation of sensor nodes with hybrid measurements in wireless sensor networks. Sensors. 2016 Jul;16(7):1143.

[2] Jaman MF, Huffman MA. The effect of urban and rural habitats and resource type on activity budgets of commensal rhesus macaques (Macaca mulatta) in Bangladesh. Primates. 2013 Jan 1;54(1):49-59.

[3] Pateman T. Rural and urban areas: comparing lives using rural/urban classifications. Regional trends. 2011 Jun;43(1):11-86.

[4] Felamban M, Shihada B, Jamshaid K. Optimal node placement in underwater wireless sensor networks. In2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA) 2013 Mar 25 (pp. 492-499). IEEE.

[5]Padwal SC, Kumar M, Balaramudu P, Jha CK. Analysis of environment changes using WSN for IOT applications. In2017 2nd International Conference for Convergence in Technology (I2CT) 2017 Apr 7 (pp. 27-32). IEEE.

[6] Zheng Y, Capra L, Wolfson O, Yang H. Urban computing: concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology (TIST). 2014 Sep 18;5(3):1-55.

[7] Borgia E. The Internet of Things vision: Key features, applications and open issues. Computer Communications. 2014 Dec 1;54:1-31.

[8] Barbosa AE, Fernandes JN, David LM. Key issues for sustainable urban stormwater management. Water research. 2012 Dec 15;46(20):6787-98.

[9] Amuthadevi, C., Vijayan, D.S. & Ramachandran, V. Development of air quality monitoring (AQM) models using different machine learning approaches. J Ambient Intell Human Comput (2021). https://doi.org/10.1007/s12652-020-02724-2

[10] Li L, Rong S, Wang R, Yu S. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review. Chemical Engineering Journal. 2021 Feb 1;405:126673.

[11] Heo S, Nam K, Tariq S, Lim JY, Park J, Yoo C. A hybrid machine learning–based multi-objective supervisory control strategy of a full-scale wastewater treatment for cost-effective and sustainable operation under varying influent conditions. Journal of Cleaner Production. 2021 Apr 1;291:125853.

[12] Sundui B, Calderon OA, Abdeldayem OM, Lázaro-Gil J, Rene ER, Sambuu U. Applications of machine learning algorithms for biological wastewater treatment: Updates and perspectives. Clean Technologies and Environmental Policy. 2021 Jan 2:1-7.

[13] Cai W, Long F, Wang Y, Liu H, Guo K. Enhancement of microbiome management by machine learning for biological wastewater treatment. Microbial Biotechnology. 2021 Jan;14(1):59-62.

[14] Gencosman BC, Sanli GE. Prediction of Polycyclic Aromatic Hydrocarbons (PAHs) Removal from Wastewater Treatment Sludge Using Machine Learning Methods. Water, Air, & Soil Pollution. 2021 Mar;232(3):1-7.

[15] Yang J, Zhou A, Han L, Li Y, Xie Y. Monitoring urban black-odorous water by using hyperspectral data and machine learning. Environmental Pollution. 2021 Jan 15;269:116166.

[16] Alvares-Sanches T, Osborne PE, White PR. Mobile surveys and machine learning can improve urban noise mapping: Beyond A-weighted measurements of exposure. Science of The Total Environment. 2021 Jun 25;775:145600.

[17] Zhao Y, Deng G, Zhang L, Di N, Jiang X, Li Z. Based investigate of beehive sound to detect air pollutants by machine learning. Ecological Informatics. 2021 Mar 1;61:101246.

[18] Chandra B, Middya AI, Roy S. Spatio-temporal prediction of noise pollution using participatory sensing. InEmerging Technologies in Data Mining and Information Security 2021 (pp. 597-607). Springer, Singapore.

[19] Van Hauwermeiren W, Filipan K, Botteldooren D, De Coensel B. Opportunistic monitoring of pavements for noise labeling and mitigation with machine learning. Transportation Research Part D: Transport and Environment. 2021 Jan 1;90:102636.

[20] Khatri, S., Vachhani, H., Shah, S. et al. Machine learning models and techniques for VANET based traffic management: Implementation issues and challenges. Peer-to-Peer Netw. Appl. 14, 1778–1805 (2021). https://doi.org/10.1007/s12083-020-00993-4

[21] Khan S, Nazir S, García-Magariño I, Hussain A. Deep learning-based urban big data fusion in smart cities: Towards traffic monitoring and flow-preserving fusion. Computers & Electrical Engineering. 2021 Jan 1;89:106906.

[22] Li C, Xu P. Application on traffic flow prediction of machine learning in intelligent transportation. Neural Computing and Applications. 2021 Jan;33(2):613-24.

[23] Joshi A, Miller C. Review of machine learning techniques for mosquito control in urban environments. Ecological Informatics. 2021 Jan 26:101241.

[24] Yang H, Liu X, Zhang D, Chen T, Li C, Huang W. Machine learning for power system protection and control. The Electricity Journal. 2021 Jan 1;34(1):106881.

[25] Alimi OA, Ouahada K, Abu-Mahfouz AM. A review of machine learning approaches to power system security and stability. IEEE Access. 2020 Jun 19;8:113512-31.

[26] Malbasa V, Zheng C, Chen PC, Popovic T, Kezunovic M. Voltage stability prediction using active machine learning. IEEE Transactions on Smart Grid. 2017 Apr 12;8(6):3117-24.

[27] Karimipour H, Dehghantanha A, Parizi RM, Choo KK, Leung H. A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids. IEEE Access. 2019 May 31;7:80778-88.

[28] Tian W, Zhu C, Sun Y, Li Z, Yin B. Energy characteristics of urban buildings: Assessment by machine learning. InBuilding Simulation 2021 Feb (Vol. 14, No. 1, pp. 179-193). Tsinghua University Press.

[29] Kim K, Park J, Lee J. Fuel Economy Improvement of Urban Buses with Development of an Eco-Drive Scoring Algorithm Using Machine Learning. Energies. 2021 Jan;14(15):4471.

[30] Zong-kui WU, Yu-feng FA. Design of toxic and harmful gas monitoring system based on machine learning method. Fire Science and Technology. 2020 Nov 15;39(11):1550.

[31] Hanga KM, Kovalchuk Y. Machine learning and multi-agent systems in oil and gas industry applications: A survey. Computer Science Review. 2019 Nov 1;34:100191.

[32] Mashechkin IV, Petrovskiy MI, Tsarev DV, Chikunov MN. Machine learning methods for detecting and monitoring extremist information on the Internet. Programming and Computer Software. 2019 May;45(3):99-115.

[33] Yuan FG, Zargar SA, Chen Q, Wang S. Machine learning for structural health monitoring: challenges and opportunities. InSensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2020 2020 Apr 23 (Vol. 11379, p. 1137903). International Society for Optics and Photonics.

[34] Kafy AA, Naim NH, Khan MH, Islam MA, Al Rakib A, Al-Faisal A, Sarker MH. Prediction of urban expansion and identifying its impacts on the degradation of agricultural land: a machine learning-based remote-sensing approach in Rajshahi, Bangladesh. InRe-envisioning Remote Sensing Applications 2021 Mar 3 (pp. 85-106). CRC Press.

[35] Ogunyemi OI, Gandhi M, Lee M, Teklehaimanot S, Daskivich LP, Hindman D, Lopez K, Taira RK. Detecting diabetic retinopathy through machine learning on electronic health record data from an urban, safety net healthcare system. JAMIA open. 2021 Jul;4(3):ooab066.

[36] Alghamdi A, Hammad M, Ugail H, Abdel-Raheem A, Muhammad K, Khalifa HS, El-Latif A, Ahmed A. Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. Multimedia tools and applications. 2020 Mar 23:1-22.

[37] Vimal S, Suresh A, Subbulakshmi P, Pradeepa S, Kaliappan M. Edge computing-based intrusion detection system for smart cities development using IoT in urban areas. Internet of things in smart Technologies for Sustainable Urban Development. 2020 Apr 29:219-37.

[38] Bangui H, Buhnova B. Recent Advances in Machine-Learning Driven Intrusion Detection in Transportation: Survey. Procedia Computer Science. 2021 Jan 1;184:877-86.

[39] Choi YH, Sadollah A, Kim JH. Improvement of Cyber-Attack Detection Accuracy from Urban Water Systems Using Extreme Learning Machine. Applied Sciences. 2020 Jan;10(22):8179.

[40] Zumwald M, Knüsel B, Bresch DN, Knutti R. Mapping urban temperature using crowd-sensing data and machine learning. Urban Climate. 2021 Jan 1;35:100739.

[41] Zumwald M, Knüsel B, Bresch DN, Knutti R. Mapping urban temperature using crowd-sensing data and machine learning. Urban Climate. 2021 Jan 1;35:100739.
[42] Zumwald M, Knüsel B, Bresch DN, Knutti R. Mapping urban temperature using crowd-sensing data and machine learning. Urban Climate. 2021 Jan 1;35:100739.
[43] Golpayegani F, Ghanadbashi S, Riad M. Urban Emergency Management using Intelligent Traffic Systems: Challenges and Future Directions. In2021 IEEE International Smart Cities Conference (ISC2) 2021 Sep 7 (pp. 1-4). IEEE.
[44] Stanford KA, McNulty MC, Schmitt JR, Eller DS, Ridgway JP, Beavis KV, Pitrak DL. Incorporating HIV Screening With COVID-19 Testing in an Urban Emergency Department During the Pandemic. JAMA Internal Medicine. 2021 Apr 12.
[45] Xiong G. Intelligent city emergency intelligence perception model based on social media big data. Journal of Ambient Intelligence and Humanized Computing. 2021 Mar 11:1-4.
Dataset
[46] https://data.world/kingcounty/vwmt-pvjw#
[47] https://data.world/codefordc/storm-surge-risk-areas
[48] https://data.world/dublin-city/34788ceb-b829-445b-a12f-cbe4df751e88
[49]https://www.kaggle.com/city-of-seattle/sea-building-energybenchmarking