DEEP AIR LEARNING PREDICTION AND FEATURE ANALYSIS OF AIR QUALITY

Rekha GS[1], Indrajeet Kumar[2], DJ Vineeth[3], Kushal Pandita[4], Abhishek M Patil[5]

Abstract- In the current situation within the world, air pollution management is sitting to be a large challenge. a lot of typically than not, academic establishments in several countries suffer from a contaminated atmosphere due to inadequate designing and improper infrastructure.Students and professors in lecture halls may tolerate health problems thanks to continually being exposed to a contaminated atmosphere. In our project, we've got engineered a price effective atmosphere display which might notice completely different waste gases like CO,CO_2,NO_2,particulate matter($PM10$/$PM2.5$/$PM1$) aboard meteoric parameters like wetness and temperature. we've got determined that very same kind of sensors for the identical gases offer completely different values though the sensitivity of sensors is appropriate, therefore we've got additionally tried to calibrate sensors exploitation machine learning techniques. We have got additionally patterned a perfect category period that a schoolroom atmosphere is taken into account healthy for a specific variety of scholars using our device.
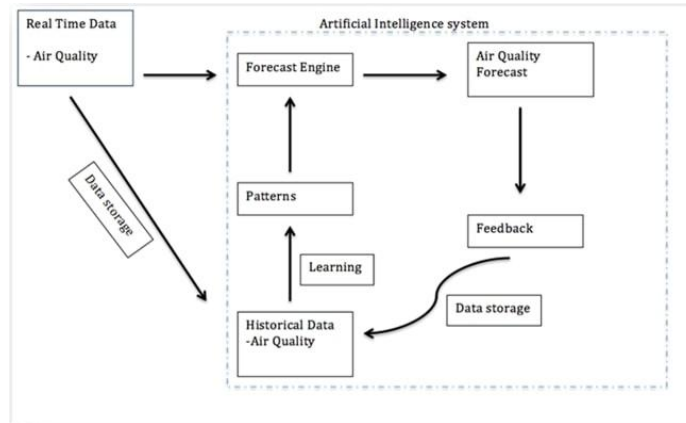**Keywords –** Machine Learning , societal impact , particulate matter, calibrate, sensors.

1. INTRODUCTION

People residing in urban areas are the foremost liable to indoor pollution. we have a tendency to cannot guarantees smart room conditions for college students unless we think about however the classroom itself is planned with relation to the ventilation and then on. Some harmful indoor pollutants embrace CO, CO2, Volatile Organic Compounds and particle matter. someone spends majority of his/her time inside, that the condition of the air within is of serious importance. Poor air quality may cause respiratory organ diseases, lesser concentration at work and even premature death. though air quality observance systems are gift to assist in these things, these tend to be bigticket and principally immovable. To alleviate this issue we've got designed a device primarily based device that is value effective, having high exactness and moveable in nature capable of measurement concentrations of most of the pollutants as mentioned on top of. The incorporation, prediction, and have analysis of air quality are 3 major disciplines within the space of urban air quality observance. An honest interpolation solves the matter that there aren't enough range of observance systems additionally, those with uneven distribution in a very city; a particular prediction offers helpful insights which might facilitate defend humans from the pollution, a significant feature analysis uncovers major factors that raise the variation of air quality. In general, the solutions to those topics will prize out valuable info to foster pollution management,and consequently generate nice techno-social impacts.
.

2. PROPOSED SYSTEM

This paper intends to deal with challenges by utilizing the knowledge contained within the data that's unlabeled and spatio-temporal, feature choice and association analysis for the urban air knowledge. Thou tagged knowledge is troublesome to get, monumental amounts of unlabeled knowledge also can be gathered cheaply. Generally, unlabeled data will facilitate in providing info to grasp the structure of knowledge higher.Following are the advantages of our proposed system: 1) Our system is compact and portable in nature. It maybe deployed in any environment, indoor or outdoor to understand the quality of weather at that location in real time. 2) Unlike the systems mentioned in the above section, our system initially classifies the Air Quality Index using a Max Operator to help ease the further processes. 3) Our aim is to develop a low cost Air Quality monitoring system boosted by machine leaning techniques.

[1,2,3,4,5] Department of Computer Science and Engineering, B.M.S College Of Engineering, Bangalore, Karnataka, India

2.1 System Architecture of Proposed system
2.1.1. Data Collection
Data collection is one methodical approach to gathering and measuring information from a variety of sources to get
a full and correct picture of an area of interest. The preliminary data collected from online sources remains in the form of plain statements . The raw data contains error, exclusions and inconsistencies. It requires corrections after careful inspecting the completed questionnaires.

2.1.2. Data Preprocessing
Data Preprocessing may be a technique that's accustomed convert the information into a clean data set. In alternative phrases, whenever the information is gathered from totally different sources it's collected in raw format that isn't possible for the analysis. Data Preprocessing is critical due to the presence of unformulated real-world knowledge. principally real world knowledge consists of -
1.Missing knowledge - There are several reasons for missing data like data isn't ceaselessly collected or a slip up in data entry and then on.
2.Noisy knowledge - the explanations for the existence of noisy data might be a technical downside of the device that collects data, a person's error throughout knowledge entry and then on.
3.Inconsistent knowledge - The presence of inconsistencies are because of the explanations such existence of duplicate entries, human knowledge entry, containing mistakes in names.

2.1.3. Working of System
The system developed has a two-fold interface according to the type of user namely, Admin and User. The admin is in charge of Training the machine learning model and to ensure the model is consistent and accurate enough to predict weather quality accurately. This is done by giving the admin options to view statistics reports of the model's training. The admin first logs into the system, and uses the pre-processed data to train the model. Further, the user an
log into the system to predict the air quality where he or she has stationed the system.
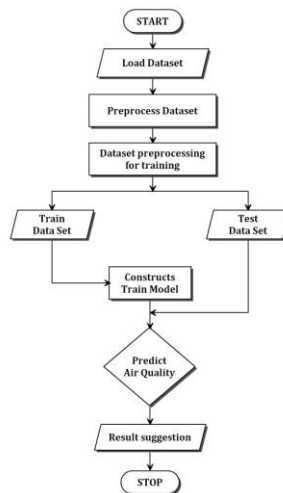
2.2 System Flowchart
Algorithms
1)Maximum-Operator: For the proposed Air Quality Index system, a maximum operator is selected :

There are primary reasons for adopting this methodology such as:

A. Free from overshadowing and equivocation.

B. Health effects of combination of pollutants are not known and thus a health based index cannot be weighted.

2) Random Forest Classifier: Random Forest is a flexible machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the widely used algorithms, owing to its simplicity and because it can be used for both classification and regression tasks.
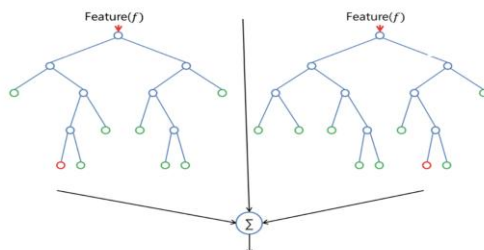

Figure 3. Random Forest as a collection of multiple decision trees.

Random Forest is a supervised learning algorithm. Like you can already see from its name, it combines a number of decision tress and somehow makes it random. The "forest" it builds, is an ensemble of Decision Trees, most of the time trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result.

4. ANALYSIS

In this final section, we are going to check our classification model on our ready dataset and additionally predict the air quality on our dataset. To guage the performance of our created classification and build it love current approaches, we have a tendency to use accuracy to live the effectiveness of classifiers. After model building, nowing the ability of model prediction on a replacement instance, is extremely vital issue. Once a prognostic model is developed victimization the historical information, one would be curious on however the model can perform on the info that it's not seen throughout the model building method. One would possibly even strive multiple model varieties for the identical prediction downside, and then, would really like to grasp that model is that the one to use for the real-world higher cognitive process state of affairs, just by examination them on their prediction performance. to live the performance of a predictor, there are ordinarily used performance metrics, like accuracy, recall etc. First, the foremost ordinarily used performance metrics are going to be delineated , then some illustrious estimation methodologies are explained and compared to every alternative. "Performance Metrics for prognostic Modeling In classification issues, the first supply of performance measurements may be a coincidence matrix.

5. ACKNOWLEDGEMENT

6. REFERENCES

[1]   Santosh Kumar Nanda, Debi Prasad Tripathi, S.S. Mahapatra, "Application of Legendre Neural Network for Air- Quality Prediction.", International Conference on Engineering and Technology, May 2011.

[2]   Surendra Rao, "Prediction of Particulate Matter Concentrations Using Artificial Neural Network." Resources and Environment 2012,2(2):30-36

[3]   Mahboudeh Afzali, Afsahen A fzali, Gholamreza Zahedi," The prediction of Artificial Neural Network Technique in daily and monthly ambient air temperature prediction" International Journal of Environmental Science and Development, Vol3, No.1, Feb 2012.

[4]   Munir Hussain Naveed, Sheikh Saeed Ahmed, Sobia Khalid and Sehresh Khan," Development of Prediction Model for the concentration level of air toxin using Artificial Neural Network." World Applied sciences journal 10(1):01- 08,2010,ISSN:1818-4952.

[5]   S. Shruti Tikhe, Dr. Mrs. K.N.Londhe " Forecasting Criteria Air Pollutants Using Data Driven Approaches; An Indian Case Study ."IOSR Journal Of Environmental Science , Toxicology and Food Technology, Vol. 3, Mar-Apr 2013,PP.01-08,ISSN:2319-2399

[6]   Suraya Ghazali,Lokman Hakim Ismail, "Air Quality Prediction Using Artificial Neural Network.",IconCEES 2011 , 3-5 April 2012 ,Johor Bahru, Malaysia

[7]   P.Zannetti,D.Al-Ajmi,.S.Al-Rashied,"An Introduction to Air Pollution- Definitions,Classification and History