

# Surveying Madrid Air Pollution

K. Subha<sup>1</sup>, Aritra Mondal<sup>2</sup>

<sup>1</sup>Assistant Professor, Dept. of Computer Science and Engg., SRM Institute of Sci. and Tech., Chennai, India

<sup>2</sup>Student, Dept. of Computer Science and Engg., SRM Institute of Sci. and Tech., Chennai, India

**Abstract:** Surveying Madrid Air Pollution is a challenging problem in deception detection, and it has tremendous real-world political and social impacts. Ozone and PM10 constitute the major concern for air quality of Madrid. This paper addresses the problem of the prediction of such two pollutants, using to this end several statistical approaches. In particular, feed-forward neural networks (FFNNs), currently recognized as state-of-the-art approach for statistical prediction of air quality, and are compared with two alternative approaches derived from machine learning: pruned neural networks (PNNs) and lazy learning (LL). PNNs constitute a parameter-parsimonious approach, based on the removal of redundant parameters from fully connected neural networks; LL, on the other hand, is a local linear prediction algorithm, which performs a local learning procedure each time a prediction is required. The global environment is currently facing a major issue of air pollution. It is one of the foremost cause of environmental and social health risks in India. Air pollution poses a serious threat to living things, eco-system and climate, especially on human health in densely populated urban areas where the pollution levels continuously starts increasing above the safer limits. Real time air quality monitoring systems requires special features like exact measurement of the parameters and analysis of the same. It makes decision making on timely basis and very easy for monitoring and controlling air quality. Currently monitoring urban air quality is critical subject that needs to be looked after for enhancing the well-being of citizens. The ultimate target of these systems is to provide exact, live information concerning the air pollution threats and to the relevant authorities for taking the necessary decisions and actions to improve the air quality. Real time representation of current scenario can be seen through such applications which allows to do health impact evaluations. The abstract comes here

**Keywords:** Natural Language Processing (NLP), Random Forest, Random keyword, AI, ML.

## 1. Problem statement

Air pollution is a tremendous problem in big cities, where health issues and traffic restrictions are continuously increasing. The concentration of Nitrogen Dioxide (NO<sub>2</sub>) is commonly used to determine the level of pollution. In Madrid, Spain, There are several stations in different parts of the city they are constantly collecting the NO<sub>2</sub> levels. A set of alerts based on the NO<sub>2</sub> level have been defined to monitor and avoid high pollution levels. The main goal of this project is to predict a pollution alert in advance. A pollution level alert means that the pollution has reached one of the set levels which is set as a bar to depict the harm level of the pollution.

Pollution is now a common place term that our ears are

attuned to. We hear about the various forms of pollution and read about it through the mass media. Air pollution is one such form that refers to the contamination of the air, irrespective of indoors or outside. A physical, biological or chemical alteration to the air in the atmosphere can be termed as pollution. It occurs when any harmful gases, dust, smoke enters into the atmosphere and makes it difficult for plants, animals and humans to survive as the air becomes dirty.

Air pollution can further be classified into two sections- Visible air pollution and invisible air pollution. Another way of looking at Air pollution could be any substance that holds the potential to hinder the atmosphere or the well-being of the living beings surviving in it. The sustainment of all things living is due to a combination of gases that collectively form the atmosphere; the imbalance caused by the increase or decrease of the percentage of these gases can be harmful for survival. The Ozone layer considered crucial for the existence of the ecosystems on the planet is depleting due to increased pollution. Global Warming, a direct result of the increased imbalance of gases in the atmosphere has come to be known as the biggest threat and challenge that the contemporary world has to overcome in a bid for survival.

## 2. Introduction

This project gives the brief theoretical preview about Machine Learning and development thus going through the essence of the problem that should be resolved for the users operating the machine learning. The project aims to develop a program which deals with the increasing rate of air pollution of Madrid and how the gases present in the air can be calculated. The amount of gases is presented with the help of percentage of the total amount of the air. Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed. Machine learning (ML) process is an approach to developing artificial intelligence. Is a method of data analysis that automates analytical model building. The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people. Instead of writing code, you feed data to the generic algorithm, and it builds logic based on the data given. I used Anaconda as it is one of the most used Integrated Development Environment (IDE) for ML. Languages used for ML are Java, C, C++ and Python. However python is the most used language out of all as it is more easy to write. Python is an

interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library. Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of Python's other implementations. Python and CPython are managed by the non-profit Python Software Foundation. It was in 1940s when the first manually operated computer system, ENIAC, was invented. At that time the word "computer" was being used as a name for a human with intensive numerical computation capabilities, so, ENIAC was called a numerical computing machine. In the 1950s, we see the first computer game program claiming to be able to beat the checkers world champion. This program helped checkers players a lot in improving their skills! Around the same time, Frank Rosenblatt invented the Perceptron which was a very, very simple classifier but when it was combined in large numbers, in a network, it became a powerful monster. Thanks to statistics, machine learning became very famous in 1990s. The intersection of computer science and statistics gave birth to probabilistic approaches in AI. This shifted the field further toward data-driven approaches. Having large-scale data available, scientists started to build intelligent systems that were able to analyze and learn from large amounts of data. As a highlight, IBM's Deep Blue system beat the world champion of chess, the grand master Garry Kasparov. My document consists of detailed description of my internship i.e. the importance of my domain current industry, skill required for creating algorithms, survey on the dream company, current projects of my dream company and the future expectations of my domain and concluding part explaining the need of domain in the current industry. The internship helped me to gain knowledge in the field of ML and helped me to update with the latest and trending technology that ever existed in the market, because the demand and value of ML is never going to fall down and with the latest updates of ML.

### 3. Modules (Methodology)

The project aims at helping people to detect the rate at which the pollution or the pollutants within the air may increase or decrease. The project allowed me to go through the depth of the algorithm creation and what are the difficulties that are faced by the developer to create a good running program with a good user interface. The creation of algorithm is a very sophisticated process so a minor mistake can lead fatal error and any lead to crashing of program and make the algorithm unable to perform

task for which it was made.

The coding part or the backend of is the complicated and very important part of program creation as it would deal with all the things that would be responsible for each and every buttons pressing and connecting it through other buttons and even connecting it to websites that would give the description of the places and Machine Learning is a cutting edge technology right now in today's current IT industry. Having the power to build and design applications is a boon. It helped me nourish my Java knowledge to a deeper extent. Although I built a simple application, but it gave me tremendous confidence to work on much bigger projects in the near future. I learnt to work on how to handle manifests and how they are to be connected. The inter working of an application and how each and every component are useful. All of them define the application. Machine Learning also made me learn how to design an app efficiently. The app must be developed in such a way so that users find it convenient and attractive. I could also learn about Python database which is so predominantly used in Machine Learning. It was a good experience learning ML and Python together.

So basically the objective of the air monitoring or the air detection system are as follows:

- Determining the level of contaminants in an airshed to compare with standards and guidelines
- Reporting on the state of the environment
- Obtaining exploratory data
- Conducting air quality research
- Obtaining data for air quality modelling
- Providing air quality information for policy or strategy development
- Assessing the effectiveness of policy based on air quality trends.

### 4. Implementation

The concept of machine learning has been getting more and more adopted in various fields of interest, with different meanings and implementations. In daily usage such as virtual personal assistants such as Siri, Alexa, Google Now. They already made things so easier for humans. Machine learning has been getting more used as it can predict more from the information and data used earlier by the user. They get to know more about the preferences and then later help to render results tailored to preferences. Starting from the analysis of existing works, I designed a scenario to identify ML based application features for both participants and exhibitors. I propose a modular architecture able to provide service for all stakeholders, it offers a baseline for exhibitors who want to provide new interactive services to participants.

Motivated by the real world AI working, propose a novel predicting mechanism to analyze and classify the data.

The concept of machine learning has been getting more and more adopted in various fields of interest, with different meanings and implementations. In daily usage such as virtual personal assistants such as Siri, Alexa, Google Now. They

already made things so easier for humans. Machine learning has been getting more used as it can predict more from the information and data used earlier by the user. They get to know more about the preferences and then later help to render results tailored to preferences.

Starting from the analysis of existing works, I designed a scenario to identify ML based application features for both participants and exhibitors. I propose a modular architecture able to provide service for all stakeholders, it offers a baseline for exhibitors who want to provide new interactive services to participants.

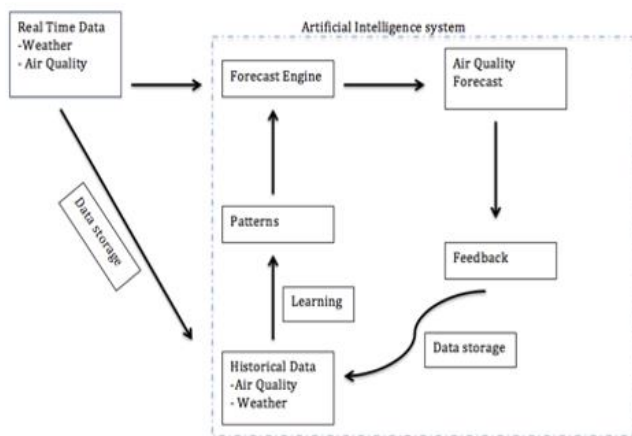


Fig. 1. Flow chart of the air quality forecast used by Airvisual

### 5. Conclusion

The benefit of any research lies in the result. The proposed system delivers excellent results. The compelled issue was the proposed system over existing system. The Mathematical calculations techniques in the proposed system as mentioned stood up picturing the drawback of existing system. Python proved excellent for implementation of mathematical calculations, their simulating results and graphical display. This tool further helped in analysis of the very two System’s scenario. The declaration reveals the necessity to change for attaining up gradation of WSN for monitoring air pollution.

A limitation of this research was that models were not developed separately by season because of the limited data record. When more data are available, it could be worth investigating using separate models for different seasons. Another limitation is the model structure of OSELM and MLP NN methods. The number of hidden nodes were decided according to the initial training data and did not change in

subsequent model updates because of limiting computational resources, resulting in xed model complexity. As new data arrive, information on larger term variability, e.g. interannual or interdecadal variability, becomes available, but the xed model complexity would not have the capacity to learn the additional structure in the data. If model complexity can be changed during the updating process, this may enhance the prediction skills.

### References

- [1] Shifeng Fang, Li Da Xu, Senior Member, IEEE, Yunqiang Zhu, Jiaerheng Ahati, Huan Pei, Jianwu Yan, and Zhihui Liu, An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things, IEEE Transactions On Industrial Informatics, Vol. 10, No. 2, May 2014, 2014 IEEE.
- [2] Chen Xiaojun, Liu Xianpeng, Xu Peng, IOT- Based Air Pollution Monitoring and Forecasting System, 2015 International Conference on Computer and Computational Sciences (ICCCS)
- [3] Ling Wang, Xi-yuan Xiao, Jian-yao Meng, Prediction of Air Pollution Based on FCM-HMM Multi-model, Proceedings of the 35th Chinese Control Conference July 27-29, 2016, Chengdu, China.
- [4] Young Jin Jung, Yang Koo Lee, Dong Gyu Lee, Keun Ho Ryu, Silvia Nittel, Air Pollution Monitoring System Based On geo sensor Network,”
- [5] Silvia Nittel, Geosensor Networks: New Challenges in Environmental Monitoring using Wireless Sensor Networks, Sensing in a Changing World, 2008.
- [6] Petra Vidnerov, Roman Neruda, Sensor Data Air Pollution Prediction by Kernel Models, 2016 16th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing
- [7] Dan S tefan Tudose, Traian Alexandru Patrascu, Andrei Voinescu, Razvan Tataroiu, Nicolae T apus, Mobile Sensors in Air Pollution Measurement.
- [8] Xiaofeng Liuly, Chaosheng Xiang, Bin Li, Aimin Jiang, Collaborative Bicycle Sensing for Air Pollution on Roadway
- [9] Shwetal Raipure, Deepak Mehete, Wireless Sensor Network Based Pollution Monitoring System in Metropolitan Cities, presented at the IEEE ICCSP 2015 conference
- [10] Luca Cagliero, Tania Cerquitelli, Silvia Chiusano, Paolo Garza, Giuseppe Ricupero, Xin Xiao, Modeling correlations among air pollution-related data through generalized association rules
- [11] Opatija, Croatia, Elena Baralis, Tania Cerquitelli, Silvia Chiusano, Paolo Garza, and Mohammad Reza Kavosifard, Analyzing air pollution on the urban environment, MIPRO 2016, May 30 - June 3, 2016.
- [12] Kavi K. Khedo, Rajiv Perseodoss and Avinash Mungur, Wsn air pollution monitoring system, International journal of wireless and mobile network
- [13] David G. Rickerby, Andreas N. Skouloudis, Big data for innovative air-pollution assessments in the era of verifiable regulatory decisions, Proceedings of the 18th Mediterranean Electrotechnical Conference MELECON 2016, Limassol, Cyprus, 18-20 April 2016
- [14] Julie Yixuan Zhu, Yu Zheng, Xiuwen Y, Victor O.K. Li, A Gaussian Bayesian Model to Identify Spatiotemporal Causalities for Air Pollution based on Urban Big Data, 2016 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs): SmartCity16: The 2nd IEEE INFOCOM Workshop on Smart Cities and Urban Computing.