THE BENEFITS OF AI TOOL APPLICATION IN THE AIR QUALITY MEASUREMENT SYSTEMS

Adis Rahmanović[1], Muzafer Saračević[2], Lejla Pezo[3]

1. Faculty of Technical Studies, University of Travnik, Travnik, Coal mine Banovići, Bosnia and Herzegovina
2. International University Novi Pazar, Serbia
3. Public Institution Secondary School Centre "Nedžad Ibrišimović" Ilijaš, Bosnia and Herzegovina

ABSTRACT

This article presents an analysis of the possibilities of applying AI tools, as well as an overview of the potential benefits of an air quality measurement system based on an Arduino microcontroller. The goal of the system is to monitor data on temperature and humidity, and then analyse the possibilities and benefits using AI tools. The data is stored in a MySQL database and displayed via a web application in the form of an interactive table with timestamps. The data is analysed using the Orange AI tool. The application system has been tested in real conditions, where it has been proven that it can reliably collect, display and analyse air quality data. This work contributes to the application of AI tools in the function of improving data analysis and taking adequate measures to improve air quality in real time.

Keywords: AI tool for data analysis, application of new technologies for air quality measurement system, application of smart solutions

INTRODUCTION

Exposure to PM particles can have serious health consequences, including respiratory problems, such as asthma, bronchitis and chronic obstructive pulmonary disease (COPD). PM particles are also associated with an increased risk of cardiovascular diseases, such as heart attack and stroke. Therefore, monitoring the concentration of PM particles in the air is vital to protect public health and prevent negative effects on the respiratory and cardiovascular systems. High air humidity can cause discomfort, the formation of mould and fungi, and aggravate allergic reactions. On the other hand, low humidity can increase susceptibility to respiratory infections, cause dry eyes, and irritation of the skin and mucous membranes. Monitoring air humidity is especially important in indoor spaces, such as homes, offices and schools, where people spend most of their time. Maintaining optimal humidity levels can improve comfort, productivity and reduce the risk of spreading diseases. Therefore, monitoring air humidity is an important aspect of maintaining a healthy and pleasant indoor environment.

Integrating these parameters into an air quality monitoring system allows for a holistic approach to protecting public health and the environment. In addition to the great impact on human life and health, high concentrations of PM particles can have a negative impact on plant life, disrupting photosynthesis and reducing plant productivity, and can also have harmful effects on animal life, especially wild animals living in urban areas or near industrial plants. This can result in a reduction in the number and diversity of species and disruption of the ecosystem. When it comes to the economic impact, increased levels of air pollution can lead to a decrease in productivity in industry and agriculture, an increase in health care costs due to an increased incidence of respiratory and cardiovascular diseases, and a decrease in tourism activities in areas with poor air quality.

The article describes new technologies and AI tools, and analyses the benefits and implementation of a smart system which can reliably measure the concentration of PM particles, air quality, and humidity, as well as temperatures. AI tools analyse the collected data from the comfort of the user's computer or mobile device. Hypothesis: There are numerous advantages to applying AI tools in air quality measurement systems.

APPLICATION OF AI TOOLS FOR AIR QUALITY DATA ANALYSIS

In the modern era, when air quality has become one of the key issues for the health and safety of the population, it is necessary to develop systems that enable automatic monitoring and analysis of environmental parameters. As part of this work, a simple air quality measurement system was developed that uses the Arduino platform to collect data on temperature, relative humidity, as well as the concentration of PM2.5 and PM10 particles. The collected data was stored in a database and used as input for advanced analysis using the Orange Data Mining tool.

Orange Data Mining is an open-source visual software package that enables the application of machine learning methods, statistical processing and interactive data visualisation.

For the purposes of the analysis, a data set with 295 rows was used, which contains the following attributes:

• Date and time of measurement
• Temperature (in degrees Celsius)
• Relative air humidity (in percentage)
• PM2.5 and PM10 concentrations (in $\mu g/m^3$)

The data was imported into Orange via the "File" widget (CSV format). For the purpose of checking the accuracy and purity of the data, the "Data Table" widget was used. Empty values and extremes (e.g., values 0.0 for PM particles) were detected, which were then processed using statistical methods:

• Empty values are replaced by mean values

• Unrealistically low values were considered as possible sensor errors and treated separately

Visual analysis was carried out using widgets such as "Box Plot", "Histogram" and "Scatter Plot". This revealed the following correlations:

• High relative humidity correlates with elevated PM10 values

• Temperature and PM2.5 show a slight negative correlation

• Oscillations of PM particle values follow daily changes (morning and evening hours)

• In periods of lower temperatures, a lower concentration of harmful particles was recorded

• The correlation between temperature and PM2.5 showed a slight negative trend

By using the k-means algorithm, data segmentation was performed into three clusters. The clustering parameters were: temperature, humidity, and PM2.5 and PM10.

• Cluster 1: low values of PM particles, dry days

• Cluster 2: mean PM values and mean humidity

• Cluster 3: high pollution values and high humidity

The "Silhouette Plot" and "Scatter Plot" widgets were used to evaluate and display clusters. These findings can be used for timely response on days with increased risk.

The goal of this phase was to develop a model that will predict the level of pollution based on the measured temperature and humidity.

The attribute "Pollution level" is classified into three classes (low, medium, high), based on the limit values for PM2.5 and PM10 according to WHO guidelines.

Applied algorithms:

• Naive Bayes,

• Decision Tree,

• Logistic regression.

The "Select Columns", "Test & Score" and "Confusion Matrix" widgets enabled performance evaluation:

• Naive Bayes: accuracy ~72%,

• Decision Tree: about 80%,

•Logistic regression: over 85% accuracy in recognising high pollution.

Orange offers a "Report" widget that automatically generates a summary of all steps with accompanying graphs. The interpretation of the findings shows:

• A clear link between high humidity and increased PM10 values;

• The cleanest air recorded on days with lower temperatures and low humidity;

• Clusters and models enable early warning and preventive action.

Orange Data Mining has proven to be an extremely useful tool in the processing and interpretation of data from IoT air monitoring systems. Its ability to quickly process large amounts of data and provide meaningful analytical insights makes it an ideal choice for projects like this.

Possible upgrades include:

• Adding sensors for additional gases (e.g., CO_2, NO_2),

• Connecting to meteorological API services,

• Implementation of real-time notifications.

In this way, the combination of ICT systems and AI tools contributes to smarter environmental risk management and public health protection.

Advantages of using the Orange tool:

• Intuitiveness and speed of work - does not require programming knowledge,

• Accessibility for researchers and students,

• The possibility of combining different analytical techniques,

• Support for all key stages of data processing – from preparation to classification and reporting.

Limitations:

• Limited number of advanced models compared to platforms like Python/Scikit-learn,

• Visualisations have limited adjustment options,

• Scalability can be an issue with large datasets.

Below is a practical example of using the Orange Data Mining tool on the data collected as part of the article, which concretely demonstrates how the theoretically described methods work in a real-case scenario.

After the data was collected using the Arduino platform and saved in .csv format, it was imported into the Orange Data Mining tool using the "File" widget. They are displayed in the "Data Table" for a basic overview, where a number of values were observed that needed to be cleaned and prepared for analysis.

In the first stage of data processing, using the "Edit Domain" and "Preprocess" widgets, the following was performed:

• Replacement of empty values with mean values (mean imputation),

• Filtration of unrealistically low values (e.g., PM values of 0.0 μg/m^3) that indicate possible sensor errors. Visual analysis was carried out through "Box Plot" and "Scatter Plot", where relations between temperature, humidity and PM values were analysed. There is a clearly observed tendency for high values of PM2.5 and PM10 to occur in conditions of high humidity and low temperature, which is also confirmed by distribution histograms.

For in-depth analysis, the k-means clustering algorithm was applied, which grouped all data into 3 clusters. Clusters were visualised using "Silhouette Plot" and "Scatter Plot", and were interpreted as:

• Cluster 1 – clear days (low PM values, low humidity),

• Cluster 2 – moderately polluted conditions,

• Cluster 3 – high pollution (especially PM10 and high humidity).

Then the classification of the state of pollution was carried out, where a new attribute "Pollution level" was created based on PM2.5 and PM10 concentrations. The attribute is categorised into:

• Low,

• Medium,

• High.

Using the "Select Columns", "Test & Score" and "Confusion Matrix" widgets, three different algorithms were applied:

1. Naive Bayes,
2. Decision Tree,
3. Logistic regression.

The results showed that logistic regression was the most successful, with an accuracy of over 85% in identifying high pollution conditions.

For example, in the actual entry from 07/12/2025, where the values were:

• Temperature: 30.5 °C,

• Humidity: 70%,

• PM2.5: 55 μg/m^3,

• PM10: 88 μg/m^3,

the system classified this entry as "high pollution", and a warning was sent to the user through the web application. The variables Temperature and PM2.5 particle concentration were selected as the input data set, while the number of clusters was pre-set to three ($k = 3$) in order to better observe the differences between measurement groups. The obtained results are shown on a scatter diagram, where each colour indicates belonging to a certain cluster:

• Cluster 1 (C1 – blue colour): This group includes measurements with low PM2.5 values (mostly close to 0 μg/m^3), at temperatures ranging from approximately 25°C to 35°C. These data indicate a period with stable and clean air,

• Cluster 2 (C2 – red colour): Higher PM2.5 values were recorded here (from about 5 to 35 μg/m^3), while temperatures are similar to those in C1. This cluster indicates periods with increased levels of air pollution, probably due to local emission sources or weather conditions that favour the retention of particles in the atmosphere,

• Cluster 3 (C3 – green colour): Contains a very small number of measurements with extremely high temperatures (around 50–60°C) and very low PM2.5 values. These values stand out from the rest of the data set and most likely represent outliers or sensor measurement errors.

This analysis enabled the rapid identification of periods with elevated concentrations of polluting particles, as well as the detection of potentially atypical data. The application of the K-Means method in Orange Data Mining proved to be useful for grouping and visual presentation of measurements, which facilitates interpretation and drawing conclusions about air quality in the observed period.

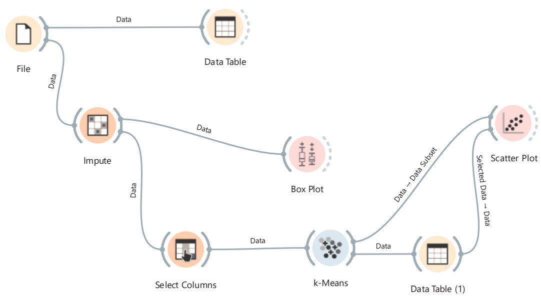

Figure 1. Visualisation of the data analysis process in the Orange Data Mining software

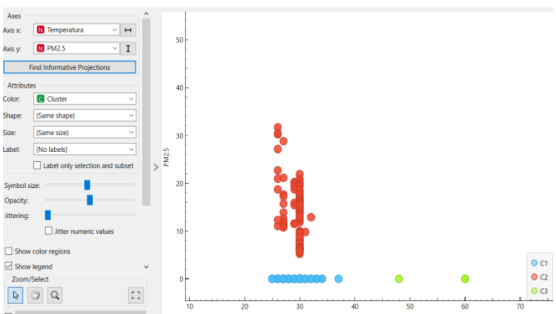

Figure 2. Visual representation of clusters in two-dimensional space (temperature – PM2.5)

The use of the K-Means algorithm in the Orange Data Mining tool brings several significant advantages:

• Simplicity and speed of data processing – the algorithm enables quick grouping of measurements and visual identification of patterns,

• Ability to recognise similar characteristics - data can be grouped according to similarities in temperature and PM2.5 concentration, which makes it easier to distinguish periods with clean air, increased pollution or unusual values,

• Visual interpretation – graphic display facilitates understanding of data and communication of results. However, the method also has certain limitations:

• It is necessary to specify the number of clusters (k) in advance - a wrong choice can lead to less precise grouping,

• Sensitivity to extreme values (outliers) – unusual measurements can affect the formation of boundaries between clusters,

• Limited number of analysed variables – only two variables were used in this analysis, so the results do not reflect all factors that affect air quality.

Despite these limitations, cluster analysis in Orange Data Mining has proven to be a useful tool for preliminary data exploration, trend detection and identification of potentially problematic time periods. The results can serve as a basis for deeper statistical processing or predictive modelling of air quality.

ANALYSIS OF BENEFITS AND POTENTIAL RISKS

The development of modern information and communication systems for monitoring air quality increasingly relies on the integration of advanced functionalities that enable far more than the basic measurement of physical parameters, such as temperature, humidity or particle concentration. The key direction of modern solutions is the installation of intelligent components that enable more efficient collection, processing and analysis of data, all with the aim of better informing users, faster decision-making and long-term environmental protection.

One of the most important characteristics of advanced systems is the ability to connect to remote servers and databases via a wireless network, which enables access to information in real time, without the need for physical presence at the measurement location. This kind of functionality enables the expansion of the system to more geographical points, which achieves wider coverage and better spatial analysis of pollution. In addition, data collected in real time can be automatically processed and analysed using data mining and artificial intelligence tools, providing deeper insights into pollution trends, possible sources and weather patterns.

In addition to accessibility and intelligent processing, modern monitoring systems must also meet high security standards. Since it is often about sensitive environmental data, which can be of public importance, it is necessary to ensure protection against unauthorised access, data manipulation and potential loss of information. Application of security protocols, communication encryption and user authentication become an integral part of the architecture of such solutions.

Another important component of advanced systems is the possibility of visualisation and a user interface that enables viewing data in a simple and understandable way. This contributes to the involvement of the wider community in monitoring environmental conditions, as even users without technical knowledge can interpret basic information and recognise potential health hazards.

In this way, such systems have both an educational and preventive function, encouraging more active participation of citizens in environmental protection.

This system enables citizens and institutions to monitor basic air quality parameters in real time, such as the concentration of PM particles, temperature and humidity. Such information can contribute to better information and timely response, such as closing windows, turning on ventilation or planning to stay outdoors in accordance with air quality. In addition, the system is scalable and can be easily expanded to a larger number of locations, which is important for cities that strive to implement “smart city” solutions. There are a number of potential possibilities for the use of AI tools in the domain of analysis of preventive as well as corrective measures to improve air quality, etc. Among the key risks is the possibility of inaccurate readings due to poor sensor calibration, especially when it comes to cheaper sensors. Also, relying on a Wi-Fi network can lead to interruptions in communication and data loss, especially if local storage of readings is not enabled. Security threats, such as attempts to gain unauthorised access to data or the device itself, should not be ignored either.

In order for the air quality monitoring system to be as reliable and resistant to potential technical and safety challenges as possible, it is necessary to apply a number of preventive and corrective measures. One of the basic techniques for increasing the precision of the data is multiple readings of the values from the sensors in certain time intervals, with the calculation of the average value. This reduces the influence of individual deviations or momentary measurement errors, which may be the result of external factors, such as changes in temperature, humidity or electromagnetic interference.

Additional reliability can be achieved by introducing local memory within the Arduino system (e.g., using a microSD module), to store data temporarily in case of internet connection loss. After re-establishing the Wi-Fi connection, data can be automatically transferred to a remote server or database, thus preventing information loss and ensuring continuity of monitoring.

In terms of security, it is recommended to apply basic protection measures that include user and device authentication when accessing the system, as well as data encryption during network transmission. In this way, the risk of unauthorised access, data manipulation or a potential cyberattack is reduced. Also, regular updating of the software part of the system (firmware and applications) enables timely removal of potential vulnerabilities and maintenance of a high level of security.

All the mentioned measures represent a framework for stable and safe functioning of the system, which is of key importance for long-term application in real conditions, especially when the system is expanded to multiple measurement points or when it becomes part of a wider IoT infrastructure in urban areas.

CONCLUSION

Through the article, we presented the application of the project, confirming the hypothesis. The system is particularly applicable in local communities, educational institutions, non-governmental organisations dealing with environmental protection, as well as for the personal needs of users who want to monitor the air quality in their environment.

The results of the AI tool application research show that this type of air quality monitoring system has a wide range of applications.

Improving the system can also include better energy management of the device, introducing solar panels for autonomous operation and increasing the range of Wi-Fi signals for remote locations. Given the increasing importance of environmental protection and digital transformation in that area, this type of system has significant potential for further development, commercialisation, and integration into larger networks for environmental monitoring (e.g., smart cities). There is a possibility of developing a mobile application that would allow users to monitor data from the device in real time, as well as the use of more advanced databases and analytical tools for long-term monitoring, statistical analysis and data visualisation.

REFERENCES

1. Arduino. (2024). Arduino Reference – Libraries. https://www.arduino.cc/reference/en/libraries/
2. Arduino official website. (2025). https://www.arduino.cc/
3. Arduino Store. (2023). Arduino Sensor Kit – Base. https://store.arduino.cc/products/arduino-sensor-kit-base
4. Orange Data Mining. (2024). Orange Data Mining Documentation. https://orangedatamining.com/docs/
5. Blum, J. (2020). Exploring Arduino: Tools and techniques for engineering wizardry. John Wiley and Sons, Inc, Canada.
6. Chen, W. (2023). Database design and implementation. ATU Faculty Open Educational Resources, 2. https://orc.library.atu.edu/atu_oer/2
7. Davila, S. Postupci i metode za procjenu i upravljanje kvalitetom zraka. https://www.researchgate.net/profile/Silvije-Davila/publication/265070893_Postupci_i_metode_za_procjenu_i_upravljanje_kvalitetom_zraka/links/5507d2860cf27e990e084649/Postupci-i-metode-za-procjenu-i-upravljanje-kvalitetom-zraka.pdf
8. Dhawan, S. (2007). Analogy of promising wireless technologies on different frequencies: Bluetooth, Wi-Fi, and WiMAX. IEEE, Sydney, Australia. https://ieeexplore.ieee.org/abstract/document/4299663/authors
9. Karan, T., Aggarwal, A., & Masih, D. (2018). Wi-Fi Technology. Scribd. https://www.scribd.com/document/373130860/WiFi-Technology-Bss
10. Komlen Lalović, I., & Živić, I. (2022). ANDROID – JAVA mobile application for presenting fingerprint scanner results in YU Info IT conference 2022, Serbia. https://www.researchgate.net/publication/359685263_JAVA_MOBILE_APPLICATION_Android_FOR_PRESENTING_FINGERPRINT_SCANNER_RESULTS
11. Krstulović, D. (2022). Application for collection and analysis of air quality data. Final paper. University of Split, University Department of Professional Studies, Department of Electrical Engineering. https://urn.nsk.hr/urn:nbn:hr:228:643448
12. Sufyan bin Uzayr. (2023). Mastering SQL. CRC Press, Boca Raton. https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.1201/9781003358435&type=googlepdf
13. Vicić, P. (2023). Wireless SMART sensors principles and applications. Final paper. Osijek. Josip Juraj Strossmayer University of Osijek. Faculty of Electrical Engineering, Computer Science and Information Technology Osijek. https://urn.nsk.hr/urn:nbn:hr:200:811068

PREDNOSTI PRIMJENE AI ALATA U SISTEEMIMA ZA MJERENJE KVALITETA ZRAKA

SAŽETAK
Ovaj rad predstavlja analizu mogućnosti primjene AI alata, ali i pregled potencijalnih prednosti kod sistema za mjerenje kvaliteta zraka baziranog na Arduino mikrokontroleru. Cilj sistema je praćenje podataka o temperaturi i vlažnosti zraka, a zatim analiza mogućnosti i prednosti pomoću AI alata. Podaci se pohranjuju u MySQL bazu podataka i prikazuju putem web aplikacije u obliku interaktivne tabele s vremenskim oznakama. Podaci se analiziraju pomoću Orange AI alata. Aplikacijski sistem je testiran u stvarnim uslovima, gdje je dokazano da može pouzdano prikupljati, prikazivati i analizirati podatke o kvaliteti zraka. Ovaj rad doprinosi primjeni AI alata u funkciji unapređenja analize podataka i poduzimanja adekvatnih mjera za unapređenje kvaliteta zraka u realnom vremenu.

Ključne riječi: AI alat za analizu podataka, primjena novih tehnologija za sistem mjerenja kvaliteta zraka, primjena pametnih rješenja

Correspondence to: Adis Rahmanović
Faculty of Technical Studies, University of Travnik, Travnik, Coal mine Banovići, Bosnia and Herzegovina
E-mail: cefes06@gmail.com