DOI: 10.5937/IOC25700S

Original scientific paper

THE PATH TO BETTER EDUCATION OF ENGINEERS USING OPEN DATA - A PRACTICAL EXAMPLE IN THE REPUBLIC OF SERBIA

Predrag Stolić[1a], Zoran Jovanović[2a], Željko Mravik[2b], Marko Jelić[2c], Sonja Jovanović[2d], Aleksandra Milosavljević[3a]

[1] University of Belgrade - Technical Faculty Bor, V.J. 12, 19210 Bor, Serbia
[2] University of Belgrade - "Vinča Institute for Nuclear Sciences - National Institute of the Republic of Serbia", Mike Petrovića Alasa 12-14, 11351 Vinča, Belgrade, Serbia
[3] Mining and Metallurgy Institute Bor, Alberta Ajnštajna 1, 19210 Bor, Serbia
[1a] pstolic@tfbor.bg.ac.rs, 0000-0002-2574-4765
[2a] zjovanovic@vin.bg.ac.rs, 0000-0003-1727-4852
[2b] mravik@vin.bg.ac.rs, 0000-0003-2327-5165
[2c] marko.jelic@vin.bg.ac.rs, 0000-0003-3723-4313
[2d] sonja.jovanovic@vin.bg.ac.rs, 0000-0003-3577-5060
[3a] alekandra.milosavljevic@irmbor.co.rs, 0000-0003-3841-7357

Abstract

This paper presents a real-world case of using open data to improve learning outcomes in the education of future graduate engineers based on the use of open data in synergy with an open API to collect real-time data. This enables students to acquire knowledge based on real-world data rather than relying on artificially generated datasets. Such an approach allows students to encounter real-life examples during their studies and to become familiar with the entire process - from data collection at the source, through storage, filtering, and other processing steps, to obtaining meaningful and usable final results.

***Keywords**: Higher Education, Octave, Open API, Open Data, Python*

1. INTRODUCTION

Today, education entails, among other things, preparing students to work in data-driven environments [1], and more recently there are growing need for using project-based learning approaches for improving students skills and better preparation for real-world problem solving [2-4]. There is also a growing need to develop competencies through appropriate AI approaches guided by relevant principles and rules [5]. Many of these involve developing own infrastructure for generating data intended for student education, as well as leveraging advancements in the field of AI, either as a standalone solution or as a complementary approach to the aforementioned methods [6].

However, these solutions are not without existing issues, nor those expected to emerge as more significant challenges in the future, affecting users across various domains, including higher education. Emerging technologies, where AI has become an indispensable component, raise new concerns regarding the reliability of data to be used in the coming years. This is largely due to the growing prevalence of synthetic data, generated by AI itself, which introduces various risks such as data pollution, bias, lack of transparency, and related issues [7–11]. Moreover, practical implementations are already encountering certain complexity-related challenges that are likely to persist for some time, particularly during transitional periods [12].

In line with the above, in the following sections of the paper, a potential solution aimed at overcoming the aforementioned challenges will be presented. This solution is based on well-established and widely adopted techniques related to the use of open data sets. It fully relies on the application of the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) in the use of digital assets [13]. The proposed solution was implemented using resources available in the Republic of Serbia and was applied in the education of students through teaching processes within the courses Process Measurement Technique and Automation of Technological Processes at the University of Belgrade - Technical Faculty in Bor [14].

2. DATA ACQUISITION

On the official Open Data Portal of the Republic of Serbia [15], a variety of open data sets can be found, which can be used in higher education to familiarize students with different forms of processing real data, whether it involves data clustering, extracting subsets, statistical analysis, or other activities. However, students are not typically introduced to the initial stages of the process, such as data acquisition and storage. This raises the question of how to leverage open data in a way that allows students to be involved in the entire process, from data generation and acquisition from source to their processing and the presentation of results at the destination. The aforementioned is achievable if there is a certain mechanism, typically in the form of an appropriate API, that enables the proper collection of data considered to be open data.

Serbian Environmental Protection Agency (SEPA) has a system for real-time air quality monitoring based on data obtained from the National Network of Automatic Air Quality Monitoring Stations [16]. Recently, an OpenData API has been made available, allowing individuals to collect hourly air quality data. Since this is a publicly available service and accessible data, there are no barriers for using it in the teaching process. It is important to note that these are hourly data from the corresponding measurement stations. The goal is purely to train students to work with real-time, real-world data. The Air Quality OpenData API represents the aforementioned appropriate API for accessing the open data system. The entire system is documented through publicly available documentation [17], as shown in Figure 1.

The data is distributed to end users in the well-known JSON format via the API. This has enabled the development of appropriate software for students, which, thanks to this API, will collect real-time data for a specified time interval. For student purposes, the interval is defined on a daily basis and can range from a minimum of 1 to a maximum of 30 days. The data collection is limited to these time intervals in accordance with the practical requirements for conducting hands-on work with students within a single semester. Therefore, the maximum period is more than sufficient for introducing students to the entire process and obtaining an adequate data set for further work.

The software is written in Python, with data storage performed locally within an appropriate Excel file. The software allows, in addition to selecting the data collection time interval, the choice of the specific measurement station from which the data will be collected, as well as which data will be gathered. The software does not require significant resources for use and is not complex; instead, it has been maximally simplified to ensure better understanding and usability by students. It is worth noting that the software has been developed to be cross-platform for computers, with corresponding versions available for Windows, macOS, and Linux platforms. Therefore, there are no obstacles to its use in terms of platform targeting. There is also the potential for developing similar software for mobile operating systems like Android, but such functionality has not yet been implemented.

Protective functions have also been implemented in the software to prevent the occurrence of certain errors. For example, there is a safeguard mechanism that prevents the same hourly value from being read and recorded multiple times. In the event of network issues, the software also includes specific procedures that attempt to retrieve the previous hourly value at ten-minute intervals. If unsuccessful before the arrival of the new hourly value, it proceeds with the new value and records an empty field for the missing one.

3. DATA PROCESSING

Data acquisition and storage are important parts, but it is essential to complete the process with further data processing and the extraction of relevant conclusions from the data. The software allows students working with data obtained in real-time and from real, dynamic environment, rather than working with generic data sets that are typically used and often fail to adequately represent the environment. An example of a code snippet used by students is shown in Figure 2.

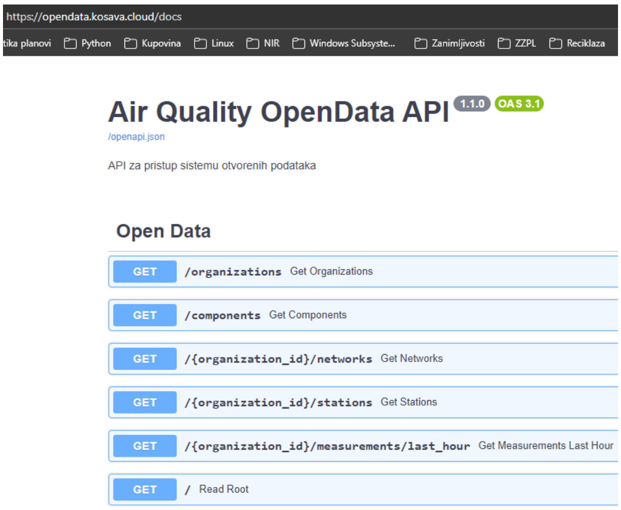

Figure 1. Official documentation for using Air Quality OpenData API

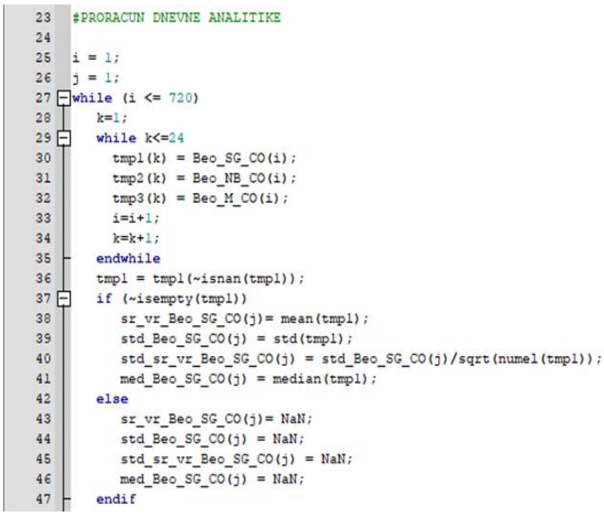

Figure 2. Part of code for data analysis written in Octave programming language.

In further work, students use the data set from the corresponding generated Excel file. In the considered case, students learn about further data processing using the Octave, which is highly suitable for use in the academic community. This is due to its math-oriented syntax and its appropriate visualization methods for both two-dimensional and multi-dimensional data representation. Additionally, in line with the aforementioned cross-platform orientation of the entire solution, Octave is also available in free versions for Windows, macOS, and Linux platforms, providing students with a certain level of platform independence.

Through this approach, students are trained to extract values from the obtained Excel file, which contains the collected data, and to format and represent these values appropriately for further manipulation and processing by the computer. Students are also taught how to appropriately handle missing values, as well as improperly represented values (e.g., in this case, the appearance of so-called Not-a-Number or NaN values), by using suitable approaches and functions (such as checks with functions like isempty, isnan, and similar). Then, students' skills are directed toward recognizing patterns in the automated data processing and making the data processing faster and more optimal in various aspects of the analysis. Regarding the calculation of specific parameters and drawing relevant conclusions from the obtained real data, students primarily perform basic statistical analysis on the data: calculating the mean, median, deviations, standard deviation, maximum and minimum values, classes, frequencies, histogram representations, and similar tasks. Finally, students are trained to record the results of the processing back into a separate Excel file locally, to maintain the consistency of the original data set with real values.

4. CONCLUSION

It is possible to train students to work in real conditions using real data without additional investments in infrastructure or implementing complex solutions. Moreover, in addition to utilizing open data access, Free and Open Source Software (FOSS) solutions were employed, meaning that the implementation of the entire system does not require significant financial investment. Additionally, given its public availability and the software licensing, which allows for free use and redistribution, there are no legal barriers to its use. In this regard, it can be said that the proposed solution is sustainable for use in the academic domain. The results of applying such a system are yet to be seen, as the entire system is currently in the pilot phase.

ACKNOWLEDGEMENTS

This research was supported by the Science Fund of the Republic of Serbia, grant No. 6706, Low-dimensional nanomaterials for energy storage and sensing applications: Innovation through synergy of action – ASPIRE *and by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (contracts No. 451-03-137/2025-03/200131, 451-03-136/2025-03/200017 and 451-03-136/2025-03/200052)*

REFERENCES

[1] T. Mucci, What is data-driven decision-making?, IBM, 2024. https://www.ibm.com/think/topics/data-driven-decision-making
[2] R. Indiwara, J. Educ. Technol., 3 (1) (2025) 85–95.
[3] F.N. Kausar, The Critical Review of Social Sciences Studies, 3 (2) (2025) 2903-2916.
[4] S. Pandey, H. Gaikwad, S. Ghawshinde, JEET, 39 (2025) 34–43.
[5] Policy guidance on AI for children: Recommendations for building AI policies and systems that uphold child rights (Version 2.0), UNICEF, 2021.
[6] P. Stolić, Z. Jovanović, S. Jovanović, Ž. Mravik, M. Grujičić, M. Jelić, Book of proceedings of the 1st International Scientific Conference Education and Artificial Intelligence (EDAI 2024), November 29-30, Vranje, Serbia, 2024.
[7] R. Kniaz, The iNcoming Tidal Wave of Data Pollution in AI, Forbes, 2023. https://www.forbes.com/sites/robkniaz/2023/05/09/the-incoming-tidal-wave-of-data-pollution-in-ai/
[8] X. Xing, F. Shi, J. Huang, et al., Nat. Mach. Intell., 7 (2025) 172–180.
[9] E. Ntoutsi, P. Fafalios, U. Gadiraju U, et al., WIREs Data Mining Knowl Discov, 10 (3) (2020) e1356.
[10] A. Arora, M. Barrett, E. Lee, E. Oborn, K. Prince, Inf. Organ., 33 (3) (2023) 100478.
[11] T. Wiehn, S&S, 22 (4), 2024, 472-476.
[12] G. Vial, J. Jiang, T. Giannelia, A. F. Cameron, MIT SMR, 62 (2) (2021) 47-53.
[13] M.Wilkinson, M. Dumontier, I. Aalbersberg, et al., Sci Data, 3 (2016) 160018.
[14] Study Program: Mining Engineering, University of Belgrade - Technical Faculty in Bor, 2023, https://www.tfbor.bg.ac.rs/mining-engineering
[15] Open Data Portal of the Republic of Serbia, https://data.gov.rs
[16] National Network of Automatic Stations for Air Quality Monitoring, Environmental Protection Agency, Ministry of Environmental Protection, Republic of Serbia, 2015.
[17] Air Quality OpenData API, https://opendata.kosava.cloud/docs