DATA SCIENCE ON ENGINEERING TRAINING

Walfredo González Hernández[1]

Abstract: Engineering education is vital for the technological development of a country, while data science occupies an important place in scientific development. First, the fundamental concepts of data science and engineering education are discussed. Secondly, the importance of data science in engineering education is explained. Finally, it proposes a system of subjects in the curriculum to train engineers with data science. The proposal should have an impact on the training of engineers who are better prepared to tackle professional problems more efficiently and effectively.

Keywords: engineering education, data science, curriculum design. Data science in engineering.

Contents

Introduction

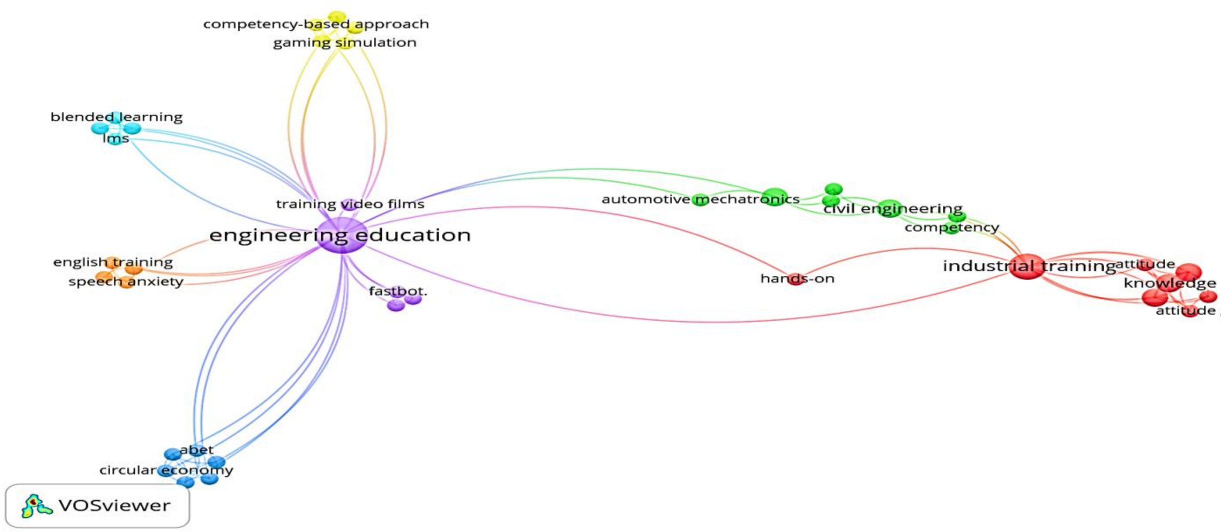
Data science is a relatively young field that brings together information science and computer science in the analysis of large amounts of data. The large volumes of data obtained in engineering processes make it necessary for professionals to dedicate themselves to its application in the different employing entities once they have graduated. Hence, their curriculum must contain the necessary elements to achieve the introduction of data processes in the shortest possible time in order to increase their efficiency and effectiveness.

The analysis of various studies related to engineering education[1-4] suggests that the introduction of data science in engineering education has not been observed. The download of

[1] Doctor in Pedagogical Sciences, Professor, Researcher; Universidad Central "Martha Abreu" de las Villas, wghernandez@uclv.cu.

articles in sciencedirect (https://www.sciencedirect.com/search?title=engineering%20training) did not yield any articles related to the subject of engineering education with data science. Figure 1 shows the occurrence of keywords in all downloaded articles and it can be seen that there is no keyword related to data science and engineering education.

Figure 1. Keyword co-occurrence map. Source: Author's elaboration.



The elements discussed so far allow us to affirm the need for research into the inclusion of data science in the engineering education curriculum. To this end, we first address the essential elements of engineering, then engineering education, followed by an argument for the need to include data science in engineering education and, finally, a proposal is made for two of the most common engineering disciplines: civil and mechanical.

Development

Engineering is the application of scientific and mathematical principles to solve problems, design, build and improve structures, machines, systems and processes. The first stage in engineering is to identify and clearly understand the problem or to define it precisely and clearly. Part of this identification is the recognition of the needs of your customers or end users in order to gather the relevant data and all necessary information about the problem. This information may include researching the problem, collecting existing data or consulting experts on the problem.

Once the data is strained, the problem is analysed by breaking it down into unwieldy parts to clarify some of the aspects of the problem. Part of this process involves using diagnostic tools and techniques to identify the root cause and potential barriers. You think of solutions and apply technical expertise and creativity to suggest different alternatives. From this point, it evaluates and selects solutions based on criteria such as feasibility, cost, impact and implementation time; taking into account a client's needs.

The next step is to design and develop the solution. To achieve this, a specific design and implementation plan is drawn up that will include prototypes, simulations and models to test the solution prior to full implementation. From this level, it carries out the necessary tests to verify that the solution works properly, validates the design, meets the objectives and respects the standards and requirements.

Implementation of the selected solution may include construction, installation and commissioning of equipment, systems or processes. This is followed by monitoring and evaluation of the performance of the solutions, with changes and improvements made to maximise efficiency and effectiveness. Finally, the entire process, including the design of prototypes, their testing and the results generated, must be documented, while communicating clearly and directly to all stakeholders the results and solutions.

Engineering education involves the use of scientific and mathematical principles to train students capable of solving problems related to the design, construction and improvement of structures, machines, systems and processes, establishing a curriculum that must be more than a compendium of elements[2] . The historical setting of engineering development, the development of the field from engineering in ancient civilisations to the present day, including historical milestones and technical advances; and the impact of the Industrial Revolution, which gave rise to the engineering profession still impacts on engineering education today. It is necessary to create inclusive environments by valuing and respecting the contributions of different cultures, integrating traditional and local knowledge specific to engineering education, and integrating and respecting cultural heritage. But an important aspect to be emphasised is the social responsibility of engineers.

The social responsibility of engineers involves not only applying technical knowledge to solve problems, but also reflecting on the consequences of one's actions on society and the environment. Employers and society in general expect engineers to provide efficient and effective solutions with ethical criteria for the benefit of society as a whole. They must develop technologies with a high degree of symbiosis with the environment, taking advantage of the resources it provides while respecting ecosystems. This deepens sustainable practices for engineering projects using recycled materials and renewable energy sources that leave a legacy of environmental cleanliness to new generations.

Engineers must comply with universal ethical principles including honesty, integrity and transparency on an ongoing basis. To achieve this, they need to ensure that their decisions are based on social and not individual well-being. Therefore, these professionals must pay attention to local development and the impact that their decisions may have on the locality. The engineer must bear the university's social responsibility as the graduate at this level who interacts with the social, local and community levels to achieve a prosperous and sustainable country.

Cultural diversity is the cornerstone on which the training of engineers must be based. To this end, they must possess a comprehensive general culture that allows them to dialogue with experts from diverse scientific, artistic and ethnic backgrounds. A general cultural vision allows for solutions that guarantee the well-being of everyone in society, with high levels of integration and that are analyzed from various perspectives. This analysis leads to training with the social responsibility that new projects demand.

Engineering students need to be trained to assess the social and environmental impacts of just and sustainable design solutions. The promotion of lifelong learning and the updating of knowledge are, on the other hand, key issues for engineers to be able to responsibly face both current and future challenges.

There are many examples of social responsibility in engineering[5] ; one example would be the design and construction of sustainable infrastructure that multiplies the use of resources and energy, reduces carbon emissions and improves the quality of life for citizens. Another example would be for engineers to work to develop clean technologies that promote clean energy and renewable energy; creating a waste management system that seeks to eliminate the environmental impact of energy sources such as wind and solar power are practices that should be adopted. Participating in community projects that provide access to clean water, sanitation and electricity in under served communities significantly improves their quality of life.

Vygotsky's cultural-historical theory is applied in the design of learning activities that foster personality development through collaborative and guided learning, backed by adequate support[6, 7] . Symbols and signs that mediate communication and enable the appropriation of new knowledge are part of this approach. Each human group in its context creates its own signs and symbols that make its language peculiar and with meanings that may escape the uninitiated[8] . Hence, communication with professionals from other areas is important as it allows diversifying and attending to the needs of these other branches of knowledge.

There are many different methodologies for engineering education, but project-based learning has become one of the most widely used[9, 10] . Projects placed using this methodology should be real so that students are confronted with integrative problems. By solving such problems, students achieve a more complete learning of the essence of engineering and its modes of action. In this process they acquire the knowledge and skills of the profession, they appropriate the language of the profession and the ways of working typical of an engineer. It is important for them to solve concrete problems under certain constraints, as is the case in organization that solve complex problems.

In this environment, students become active participants in their training process and are involved in the training processes. They access international databases, professional communities and build their learning spaces according to the profession[11] . Work on real projects integrates them into communities of professionals from whom they learn all the elements addressed in the previous paragraph, while at the same time appreciating the situations in which they can apply the knowledge and modes of action they have been trained in during their professionalization process.

Engineering problems are generally integrated with different areas of human activity, which is why its projects are inter, intra and multidisciplinary. A training process that takes these processes into account allows students to be trained in increasingly collaborative environments where respect for others prevails. It also allows them to approach from multiple angles and to approach different learning situations from the positions of other disciplines.

Mathematics plays an essential role in engineering processes. This science allows modelling and explaining in a precise and brief way the main structures and/or processes to be transformed during the engineer's performance. Another element of personality that is developed during project-based learning is the social responsibility of the students. By participating in this type of projects, students can assess the social impact of their productions, the improvements they introduce in society and how society evolves from their solutions. In these terms, they learn to value the sustainability and responsibility of their work. They become professionals committed to a prosperous and sustainable society. All the processes explained above are based on solid data collection.

Data science is based on the analysis of data in order to derive valuable knowledge from a set of data. It is achieved by combining several scientific disciplines: programming, statistics and professional skills. The process starts with data collection, drawing data from various sources such as databases, APIs, sensors and files. These are integrated to create a consistent and useful datasets. Then comes data preparation and cleaning, which includes cleaning and transforming the data to remove errors, null values and duplicate entries, and standardizing the data into a common, compatible format to facilitate analysis.

The next step is exploratory data analysis (EDA), which incorporates visual tools as well as descriptive statistics to make the distributions, relationships and patterns of data understandable[12] . In this sense, graphs and other visual aids are available to help relate patterns and behavior of the data, along with statistical methods to describe and summarize under the model to predict and/or make informed decisions. Once the model has been developed, it is analyzed for production and evaluation and machine learning models are put into production for use in real-world applications. The conditions of the models are monitored to continuously adjust those models to ensure the effectiveness of their use. In this way, data analysis is linked to informed findings, whereby the results are presented in a clear and informative way to stakeholders to enable them to understand the results in the research framework through reports, tables, presentations, and so on.

Introducing data science in the field of engineering education can help students to prepare themselves to face present and future challenges, being important that they can do data analytics projects in their studies and use predictive analytics techniques to predict behavior and trends in topics closely related to the different branches of engineering. Cases of their introduction can be predictive maintenance in mechanical engineering or with housing demand forecasting in civil engineering. Optimization algorithms are also used to improve the organization and efficiency of industrial, energy or transport systems and processes.

There are several actions that can be implemented to establish interdisciplinary collaborative relationships between engineering students and students from other disciplines in which data analysis can be relevant in the solution of complex and multifaceted problems[13] . One of them can be to share their results in various outreach spaces where the importance of data science in solving engineering problems is discussed. This participation promotes the exchange with various professionals from whom he/she can draw on and take advantage of their experience in the form of tacit knowledge.

The introduction of content associated with data science in engineering prepares them to face the various problems faced by today's engineers. The curriculum should include courses that cover essential concepts in statistics, data programming (preferably Python and R) and relational database management systems (SQL versions). The courses would enable students to understand the processes from the capture of data to its use in the solution of engineering problems. Central to this is the introduction of statistics to perform analyses using hypothesis testing. At the same time, they can develop algorithms to optimize different processes that can be automatic.

Including laboratory classes is one of the important forms of organization because it allows the student to gain the skills in analyzing numerous data to make inferences from them and obtain predictive models. These predictive models must be modeled in controlled environments before moving on to business environments. In the laboratory classes the student must solve real problems that are presented from the beginning. Hence the importance of these classes because this is where the real projects that the students must complete are solved.

In this, it is essential to take into account the tension between the complexity of the real projects to be solved, the time available, the student's possibilities and the objectives pursued by the study plan. There must be a balance between these elements, as the imbalance of one of them leads to problems in the training of engineers. Hence, the implementation of real projects must be the result of a synergy between the employing organizations where the students' internship takes place and the university. Each of these two organizations have different dynamics and must find a balance point where the most important thing is the management of the students' data science training processes.

Integrating data science into engineering courses not only helps students to better solve global problems, but also contributes to societal progress. Training future engineers with advanced skills in data analysis and modelling will drive innovation and productivity in areas ranging from manufacturing to resource management. It also fosters a culture of knowledge and data-driven decision-making that is fundamental to development and social well-being.

All the tools and topics suggested in the engineering courses are very useful for solving problems in combination with data science; above all, real engineering problems, social development and development of students' engineering skills. Through a combination of theoretical foundations, practical experience and holistic work, students not only acquire technical skills, but also develop the critical skills and ethical qualities necessary to respond to the challenges of the 21st century.

The use of special tools and programmer is another important aspect. Students should be familiar with tools such as Tableau and Power BI and big data such as Hadoop and Spark, and use programming environments such as Jupyter, Notebooks and RStudio to analyse and visualize data. It is also important to develop the right skills and ethics. It is important to be able to write effective papers and presentations that connect clearly and quickly with both technical and non-technical audiences. In addition, students will learn about the legal implications of data management, including privacy, security and access to data.

Data Science is a powerful tool for solving engineering problems, giving engineers the skills they need to analyse and use complex data. In this way, informed and affordable solutions are offered. Participation in professional education and practice is essential for engineers to meet the challenges of the 21st century.

As a goal for implementing data science and engineering education, a curriculum should be structured that emphasizes the design and organization of learning based on real projects using big data. The theory emphasizes the importance of clearly defining learning objectives, selecting appropriate materials, organizing learning experiences and evaluating the learning process to ensure that students acquire the necessary competencies.

First of all, the social work of engineers is the main objective of education. According to the philosophy of curriculum planning, it is important to set objectives that cover not only technical, but also behavioral and social relations. By applying the learning of professional ethics and social work to the project, students learn the importance of developing sound systems and sustainable solutions. This course should teach future engineers how to analyse the social and environmental impact of projects and make informed and sound decisions. In addition, it is important to encourage participation in community projects and activities that benefit society helps students develop social awareness and responsibility for the welfare of society.

In-depth knowledge of engineering and data science is another cornerstone of the curriculum. Curriculum design theory emphasizes the importance of selecting content that is relevant and current[14] . By taking courses such as data science fundamentals, machine learning and predictive modelling, students can build a solid foundation in statistics, programming and advanced data analysis techniques. These disciplines enable students to develop predictive models that optimize processes, improve efficiency and facilitate decision-making in various engineering domains. By integrating these subjects into the curriculum, students will be equipped with the technical skills needed to solve the complex challenges of modern engineering. It is important from the curricular design to elaborate proposals for the insertion of the subjects in the curriculum in engineering education. For this reason, the following curricular proposal is made for engineering degrees and later the proposal will be shown for the two engineering degrees par excellence: mechanical and civil.

Curriculum proposal to introduce data science in engineering education

- **First year**: Introduction to engineering, basic mathematics and physics, data science fundamentals course.
- **Second year**: Applied statistics, Python/R programming, data structures and algorithms.
- **Year 3**: Machine learning, advanced data analysis, practical data analysis project.
- **Fourth year**: Optimized systems, interdisciplinary projects, ethics in data science and research workshops.
- **Year 5**: Completion of studies in the project, staging in organizations with big data production and dissemination of results.

Integrated Curriculum Proposal for Civil Engineering

Year 1:

- Introduction to Civil Engineering
- Data Science Fundamentals

Year 2:

- Statistics Applied to Civil Engineering
- Programming Fundamentals (Python/R)

Third Year:

- Machine Learning for Civil Engineering
- Practical Data Analysis Project (Failure Analysis in Infrastructure)

Fourth Year:

- Advanced Data Analysis for Civil Projects
- Ethics and Responsibility in Civil Engineering
- Research and Development Seminars with Data Analysis

Fifth Year:

- Integrating Final Project (Application of Data Science in a Real Project)
- Publication of Research Results

Integrated Curriculum Proposal for Mechanical Engineering

Year 1:

- Introduction to Mechanical Engineering

- Data Science Fundamentals
- Technical Drawing and CAD

Year 2:

- Statistics Applied to Mechanical Engineering
- Programming Fundamentals (Python/R)

Third Year:

- Machine Learning for Mechanical Engineering
- Data Analysis Practical Project (Optimization of Mechanical Systems)

Fourth Year:

- Systems Control
- Advanced Data Analysis for Mechanical Systems
- Research and Development Seminars

Fifth Year:

- Integrating Final Project (Application of Data Science in a Real Project)
- Internships in Mechanical and Manufacturing Industries
- Publication of Research Results

Conclusions

Engineering education needs data science as one of its essential processes. This assertion is based on the way engineering works and thinks when implementing solutions in real environments with constraints. In this context, engineering will play an increasingly important role in the well-being and development of society as a whole.

Analyses of the introduction of data science in the educational processes of future engineers must take into account the integration of the three components of the training process: academic, work and research. Each of these processes makes a different use of data science, but they are integrated during the training processes that result in the solution of real problems in employing organizations. The last actions in the curricular proposal aim at the integration of the academic, work and research components in the communication of the solutions found to real projects.

The proposed curriculum design is based on the elements addressed in the curriculum design and the importance of data science in engineering education. The inclusion of tools from the early years is justified by the need for training in the fundamental concepts of data science and their subsequent application in the subjects of the speciality.

Bibliography

[1] S. Nocera, S. Romano, R. Francese, and G. Scanniello, "Software engineering education: Results from a training intervention based on SonarCloud when developing web apps," *Journal of Systems and Software,* vol. 222, p. 112308, 2025/04/01/ 2025, doi: https://doi.org/10.1016/j.jss.2024.112308 .

[2] V. C. Poschauko, E. Kreuzer, M. Hirz, and C. Pacher, "Engineering Education goes Lifelong Learning: Modularized Technical Vocational Education and Training Program for the Automotive Sector," *Procedia Computer Science,* vol. 232, pp. 1799-1808, 2024/01/01/ 2024, doi: https://doi.org/10.1016/j.procs.2024.02.002 .

[3] L. Rendón-Castrillón, M. Ramírez-Carmona, and C. Ocampo-López, "Training strategies from the undergraduate degree in chemical engineering focused on bioprocesses using PBL in the last decade," *Education for Chemical Engineers,* vol. 44, pp. 104-116, 2023/07/01/ 2023, doi: https://doi.org/10.1016/j.ece.2023.05.008 .

[4] V. V. Kumar, D. Carberry, C. Beenfeldt, M. P. Andersson, S. S. Mansouri, and F. Gallucci, "Virtual reality in chemical and biochemical engineering education and training," *Education for Chemical Engineers,* vol. 36, pp. 143-153, 2021/07/01/ 2021, doi: https://doi.org/10.1016/j.ece.2021.05.002 .

[5] L. M. Bland, "Examining The Quality Of Teacher-Selected Social Studies Performance-Based Assessments: A Responsive Evaluation," College of William and Mary, 2022. [Online]. Available: http://oatd.org/oatd/record?record=oai \:scholarworks.wm.edu\:etd-7306&q=title%3A%28teacher%20AND%20performance%20AND%20evaluation%29

[6] Y. Wang and Y. Ma, "Beyond Physical Space: Using Digital Technology to Support a Collaborative Cultural-Historical Educational Experiment," in *Cultural-historical Digital Methodology in Early Childhood Settings: In Times of Change, Innovation and Resilience*: Springer, 2024, pp. 39-48.

[7] N. Veresov and N. Veraksa, "Introducing cultural-historical genetic-analytical model for studying child's play: the matryoshka principle - situations within situations covered by situations," *Early Years,* pp. 1-13, 2024, doi: 10.1080/09575146.2024.2365180.

[8] W. González Hernández, M. P. Roldán, and M. M. García, "Creative learning in final year students in computer engineering: A case study of the University of Matanzas," *Thinking Skills and Creativity,* vol. 52, pp. 1-13, 2024, doi: 10.1016/j.tsc.2024.101479.

[9] A. Shekh-Abed, "Metacognitive self-knowledge and cognitive skills in project-based learning of high school electronics students," *European Journal of Engineering Education,* pp. 1-16, 2024, doi: 10.1080/03043797.2024.2374479.

[10] V. Janštová, E. Kotrčová, and T. Matějková, "Key aspects of successful science fair projects," *International Journal of Science Education,* pp. 1-17, 2024, doi: 10.1080/09500693.2024.2364956.

[11] W. González Hernández, "Los espacios de aprendizaje y las formas de organización de la enseñanza: una caracterización desde la subjetividad," *Revista de Estudios y Experiencias en Educación (REXE),* vol. 20, no. 42, pp. 17-27, 2021, doi: 10.21703/rexe.20212042gonzalez18.

[12] B. Zong, J. Li, T. Yuan, J. Wang, and R. Yuan, "Recent progress on machine learning with limited materials data: Using tools from data science and domain knowledge," *Journal of Materiomics,* 2024/07/26/ 2024, doi: https://doi.org/10.1016/j.jmat.2024.07.002 .

[13] D. Bzdok, A. Thieme, O. Levkovskyy, P. Wren, T. Ray, and S. Reddy, "Data science opportunities of large language models for neuroscience and biomedicine," *Neuron,* vol. 112, no. 5, pp. 698-717, 2024/03/06/ 2024, doi: https://doi.org/10.1016/j.neuron.2024.01.016 .

[14] J. Southworth *et al.*, "Developing a model for AI Across the curriculum: Transforming the higher education landscape via innovation in AI literacy," *Computers and Education: Artificial Intelligence,* vol. 4, p. 100127, 2023/01/01/ 2023, doi: https://doi.org/10.1016/j.caeai.2023.100127 .