SEFI Journal of Engineering Education Advancement
Vol. 1, No. 1, 2024, pp. 14-25, ISSN 3006-6301
DOI : 10.62492/sefijeea.v1i1.15

A Causation-driven Approach to Engineering Education Using Data Analytics and Machine Learning Tools

Daniela Galatro[1], Kai Hashimoto, Satya Sathwik Juttada

University of Toronto, Toronto, Canada

ABSTRACT

This work presents a causation-driven approach to engineering education using data analytics and machine learning tools as an alternative to teaching/learning approaches based on combining the physical description of the phenomenon and using first principle and/or correlations/heuristics. We aim to increase the percentage of student participation during lectures, recognize the importance of using data analytics to understand a phenomenon, introduce students to machine learning tools as complementary analysis methods, and encourage students to perform research and find "different ways" to assess and solve an engineering problem. We illustrated our approach with three examples in two different domains: lectures and supervision, including (i) understanding the physical significance of the Nusselt number using exploratory data analysis, (ii) studying the impact of the reaction temperature on conversion using Bayesian structural time-series model to evaluate the effect of an intervention, and (iii) performing a heterogeneous treatment effect to assess the potential causation of combined demographic and environmental variables on health outcomes using Causal Random Forest. To evaluate the effectiveness of our approach, we quantified the percentage of student participation, which increased by more than 20%, as well as a set of generated lessons learned that attest to deepening the acquired knowledge.

KEYWORDS

Causation, data analytics, machine learning, heterogeneous treatment effect

PUBLICATION

Submitted:
13th April 2024

Accepted after revision:
6th June 2024

Introduction

Education 4.0 (E.D. 4.0) emerged as the approach to learning that aligns with Industry 4.0 (I.D. 4.0) (Chakraborty et al., 2023). E.D. 4.0 is supported by the basic idea of using technology-based tools to deliver education while encouraging students to pursue non-traditional reasoning, fostering creativity and flexibility (Treviño-Elizondo and García-Reyes, 2022). The role of E.D. 4.0 in engineering education is crucial, as it provides the framework for developing the skills and competencies that students require for I.D. 4.0 (Chakraborty et al., 2023; Galatro et al., 2022; Treviño-Elizondo and García-Reyes, 2022). In their effort to map E.D. 4.0 and I.D. 4.0, Chakraborty et al. (2023) conducted

[1] Corresponding Author : Daniela Galatro – daniela.galatro@utoronto.ca

a bibliometric analysis to identify several components of E.D. 4.0, from which machine learning (ML) reported significant link strength.

Machine learning has been implemented in engineering education in several supportive education contexts, including prediction (academic performance, student placement), academic (decision, guidance), automatic analysis, automatic scoring, ethics, and student collaboration, among others (Sasmita and Mulyanti, 2020). ML can also be used through applications as virtual assistants, contributing to personalized and accessible learning (Machine Learning - Advanced Techniques and Emerging Applications, 2018). As a technical skill to be learned and/or an alternative approach to teaching and learning, machine learning models may support student learning by performing different data-driven tests to assess time, accuracy, and transfer learning or information migration (Liu et al., 2018) that ultimately may enhance student understanding. This enhancement can be achieved by (i) using an efficient model that allows studying the relationship (Liu et al., 2018) between variables and, consequently, (ii) simulating scenarios using this model. While it seems that "prediction" tops the capability feature of modelling using ML methods, and this may be justified over traditional process/system simulation using software to solve governing equations (first principles) to predict process/system outcomes, as ML models, categorized as black-box models, may be suited to reproduce complex systems, formerly limited by process/system simulations.

Engineering phenomena (EP) are represented by mathematical models in which there is a relationship between variables. To study EP, we primarily require (i) a mathematical description of the connection between the variables, (ii) a solution to obtain the outcome/response/output, which is compared against the measured output, (iii) defining the limitations or range of applicability of the model, and (iv) ensuring adaptability/enhancement of the model as new data is collected (generalization). Hence, mathematical models help us describe and understand EP, as we implicitly assume that causation is rationale for the relationship explaining the phenomena. Traditional ML-based modelling is known as an excellent option for accurate prediction, while lacking underlying physical meaning and/or intrinsic causation to explain the relationship between variables.

Approaches to causation include counterfactual theories, regularity theories, probabilistic theories, causal modelling, etc (Beebee et al., 2010). Causal models mathematically represent causal relationships between variables. In terms of ML-based modelling, the causation perspective allows ML to formalize a structured causal model that can reflect the effects of changes. This structure establishes cause-and-effect relationships between variables and provides insights that help us understand the phenomenon represented (Beebee et al., 2010).

Several causal ML algorithms have been developed, including tree-based algorithms, meta-learner algorithms, neural network-based algorithms, value optimization methods, and traditional methods (Zhao and Liu, 2023). Some examples of fields where they have been employed include medicine-healthcare (Prosperi et al., 2020), social sciences (Russo et al., 2019), and economics (Bermeo et al., 2023). To our knowledge, no causation-driven approach has been developed for and integrated into engineering education as a teaching/learning strategy. We hypothesize that such an approach will positively impact engineering students' learning as we expect it to contribute, as an exploratory work, to closing the gap between E.D. 4.0 and I.D. 4.0, by incorporating ML methods to represent phenomena accurately. At the same time, we look at applying the enhanced versions of these methods to support students in describing and understanding these phenomena through causation.

In this work, we present a causation-driven approach to engineering education using data analytics and machine learning tools as an alternative to traditional teaching/learning approaches based on combining the physical description of the phenomenon and using first principle and correlations/heuristics. Our approach was applied in two different domains: lectures and supervision and is illustrated with three different examples using exploratory data analysis, a causation model for time-series, and a tree-based model for estimating causation between risk factors and health outcomes.

Methodology

Our causation-driven approach is supported by reaching learning outcomes with an alternative teaching/learning approach to the one focused on the physical description of the phenomenon and using first principle and/or correlations/heuristics. It includes the use of datasets so the students can perform Exploratory Data Analysis (EDA) and machine learning tools to assess not only the relationship between variables but also causation, from which we hypothesize that it contributes to maximizing the understanding of the phenomenon while using modern data analysis/machine learning techniques that prepare our workforce.

The goals of applying a causation-driven approach to engineering education are to increase the percentage of student participation during lectures, recognize the importance of using data analytics to understand a phenomenon, introduce students to machine learning tools as complementary learning methods, and encourage students to perform research and find "different ways" to assess and solve an engineering problem. In addition to quantifying the percentage of student participation, to assess the effectiveness of our approach, we rely on continuous student feedback and acquired knowledge through presentations.

In this work, we illustrate our causation-driven approach to engineering education through three different examples, which fall into the learning domains of "undergraduate course" and "summer project". We selected these examples to illustrate its potential effectiveness in improving engineering education when teaching core undergraduate courses (heat and mass transfer), optional courses (data-based modelling), and when supervising students performing summer projects. Hence, a key element of implementing our approach is the diversification of learning domains and integration in several undergraduate courses.

In our first example, students from the second-year heat and mass transfer undergraduate course are expected to understand the physical significance of the Nusselt (Nu) number using EDA. They are provided with a dataset including the variables associated with this number, and EDA, a well-known method used in data analytics, is performed live during the lecture by running a given code developed by the instructor with the software R.

When performing EDA, students are guided by the instructor through instructions, background, and a set of related questions inducing their learning to further understand the relationship between variables and, hence, the physical significance of Nu in heat convection. A group discussion is encouraged, and the instructor requests continuous feedback to assess the effectiveness of this approach in reaching the learning outcome.

Note: The group/cohort for our Nusselt example (125 second-year students) was not familiarized with R but with other programming tools such as Python and MATLAB®. The R code and dataset were prepared by the instructor and provided to the students. Then, the instructor guided the students to run the code quickly, emphasizing the interpretation of the correlograms over the programming aspects of the exercise. The class instructor previously introduced the concept and physical meaning of the Nusselt number; hence, this guided machine learning exercise is meant to support the first principle learning process, aiming at showing alignment with the correlations. The students had no previous exposure to alternative approaches like the one presented in this work.

In our second example, we study the impact of the reaction temperature on conversion in a data-based modelling undergraduate course. The causation routine in R is run live during the lecture (by students), to analyse an intervention in time series that may explain the cause of the conversion increment as the temperature increases. The instructor provides the background information and interpretation guidelines of the selected machine learning method. At the same time, a group discussion is encouraged to compare the traditional kinetics approach (based on rate expression) and the data-driven approach while also requesting continuous feedback to assess the effectiveness of this approach in reaching the learning outcome for this task.

Note: The group/cohort for our reaction example (15 fourth-year students) was fully familiarized with R The R code and dataset were prepared by the instructor and provided to the students. Then, the instructor guided the students to run the code quickly, emphasizing the interpretation of the time-series causation (intervention) over the programming aspects of the exercise, while students were allowed to perform their own input modifications. The students had no previous exposure to alternative approaches like the one presented in this work.

In our third example, two summer students were asked to perform a heterogeneous treatment effect (HTE) analysis to assess the potential causation of combined demographic and environmental variables on health outcomes. This example arises from a project in collaboration with other disciplines and laboratories in our university, and students are expected to be exposed to HTE analysis for their first time. No prescribed method was suggested to the students, so they performed research and developed their codes in R to explain the potential causation. The supervisor required a final presentation, where a group of specialists and students engaged in discussions on the selected methods and results.

Nusselt number

The convection heat transfer mechanism is traditionally taught follows a general physical description, the analysis of the velocity and thermal boundary layers, the definition of the dimensionless Reynolds (Re), Prandtl (Pr), and Nusselt (Nu) numbers, including a discussion of their physical meaning, and finally, the derivation of the convection equations based on mass, energy, and momentum conservation and the obtention of functional forms of the Nusselt number.

The Nusselt number represents the ratio of total heat transfer due to conduction and convection to conductive heat transfer across a boundary. It is derived from a dimensionless analysis of Fourier's law.

A causation-driven approach to discussing the Nusselt number presents the non-dimensionalization of the heat transfer coefficient h with Nu and the power-law relation form of Nu, as a function of Re and Pr. This is followed by EDA between the dependent variables of Re, Pr, and hence, Nu, to analyse the relationship between variables visually.

$$Nu = \frac{h\,L_c}{k} \qquad \textit{Eq. (1)}$$

where k is the thermal conductivity of the fluid and Lc is the characteristics length.

$$Nu = C\,Re^m\,Pr^n \qquad \textit{Eq. (2)}$$

where m and n are constant exponents, and C is a geometry-dependant constant.

The Reynolds number is defined as

$$Re = \frac{\rho\,d\,v}{\mu} \qquad \textit{Eq. (3)}$$

Where ρ is the density, d is the pipe diameter, v is the velocity, and μ is viscosity.

The Prandtl number is defined as

$$Pr = \frac{Cp\ \mu}{k} \qquad Eq.\ (4)$$

Where Cp is the heat capacity and k is the thermal conductivity.

The dataset used for this example was obtained from simulating the Nusselt number using the Colburn equation for fully developed turbulent flow in smooth tubes (Colburn, 1964):

$$Nu = 0.023\ Re^{0.8}\ Pr^{1/3} \quad \begin{pmatrix} 0.7\ \leq Pr \leq 160 \\ Re\ > 10{,}000 \end{pmatrix} \qquad Eq.\ (5)$$

Gaussian noise was added to the noise-free calculated Nu, to simulate "measured" data. The corresponding ranges of the simulated data are shown in Table 1.

Table 1: Ranges of the simulated data

Variable	**Range**	**Unit**
Viscosity (Visc)	0.001 – 0.005	kg/m.s
Heat capacity (cp)	4,200 – 10,000	J/kg.K
Thermal conductivity (Cond)	0.1 – 0.9	W/m.K
Pipe diameter (d)	0.010 – 0.035	m
Density (Den)	400 – 1,000	kg/m^3
Velocity (Vel)	0.5 – 4	m/s

Note: The data generation process for this exercise requires an intermediate level of expertise in statistics/machine learning and knowledge of heat transfer. The estimated generation time (sourcing, gathering, and exploratory data analysis) was two hours.

A simple chart to show the relationship between each pair of variables is the correlogram, which assumes that such a relationship is linear. When represented by circles, their radii range between -1 and 1. Colour intensity (blue as positive and red as negative correlation) and the size/radius of the circle indicate proportionality to the correlation coefficients. While correlograms are not causation tools, they can be employed as effective visualization EDA techniques when studying known first principle phenomena and/or sound-derived relationships from dimensionless analyses. In this work, we used a correlogram to observe the relationship between the variables associated with Re (density, pipe diameter, velocity, and viscosity), Pr (heat capacity, viscosity, and thermal conductivity), and Nu (Re, Pr). The students are expected to (i) "prove" the definitions of Re, Pr, and Nu (from Eq. (2)) by qualitatively analysing the proportionality of variables and (ii) "ranking" the importance of variables by quantitatively analysing proportionality of variables. As an alternative to the traditional teaching approach to convection-Nusselt, this approach conserves the same learning outcome, which states that at the end of the lecture, students will be able to understand the physical significance of the Nusselt number.

Impact of the Reaction Temperature on Conversion

In engineering reaction and kinetics, it has been observed that, in many cases, as the temperature increases, the conversion increases first (up to complete consumption), speeding up the reaction. This phenomenon can be easily explained by analysing the reaction rate expression. As this rate or speed of reaction represents the change of concentration by the time interval, the conversion can be defined as a time-series (TS). A temperature change and its impact on reaction conversion shall be visually observed in the TS, and the kinetics principles can phenomenologically explain the causation. An alternative approach to evaluate this impact and prove causation would be using a TS dataset of reaction temperature and conversion and estimating the causal effect of the designed intervention (temperature increment). For instance, a Bayesian structural TS model can be built to predict the counterfactual from which the causal inference is performed. To illustrate our causation-driven approach to the impact of the reaction temperature on conversion, we used the function Causal Impact (in R) for a dataset that included reaction temperature and conversion TS. This tool can be particularly useful when monitoring real-time variables in chemical process plants.

The students are expected to "prove" causation between reaction temperature and conversion using the Bayesian structural TS model and cross validating the corresponding reaction rate expression. As an alternative to the traditional teaching approach to the relationship between reaction temperature and conversion (provided in the previous courses of kinetics and reaction engineering), this approach conserves the same learning outcome, which states that at the end of the lecture, students will be able to understand the effect of the temperature on the reaction rate.

Note: The data generation process for this exercise requires an intermediate level of expertise in statistics/machine learning and knowledge of reaction engineering. The estimated generation time (sourcing, gathering, and exploratory data analysis) was three hours.

Effect of the combined demographic variables and pollution exposure on health outcome

In this example, we are interested in assessing the potential causation between demographic variables, exposure to a chemical pollutant (as treatment variable, measured as concentration) and the expected probability of a health outcome. The data for this example was generated for teaching/learning purposes, from a simulated unconditional logistic regression, which are typically employed to fit data for age-matched case-control data in medicine (Kuo et al., 2018).

To observe the effect of the combined demographic variables and the pollutant exposure on the expected health outcome, our students used Causal Random Forest (CRF). CRF, is a machine learning tool that modifies the Random Forest method to estimate causal effects by including causal reasoning through splitting data to maximize differences in the relationship between an outcome and a treatment variable. CRF has been successfully used to estimate HTE in personalized medicine, marketing, environmental studies, and business decision-making problems. The average treatment effect (ATE) is used as a causation metric for the CRF method, measuring the difference in mean outcomes between individuals assigned to the treatment and individuals assigned to the control. ATE values different than zero suggest causation.

Note: The data generation process for this exercise requires an advanced level of expertise in statistics/machine learning. The estimated generation time (sourcing, gathering, and exploratory data analysis) was one week.

Results

This section summarizes the main results and findings of implementing our causation-driven approach to engineering education using data analytics and machine learning tools through three examples falling into two different learning domains: undergraduate courses and summer projects.

Nusselt number

Figure 1 shows a correlogram for the relationship between variables affecting the Nusselt number. Through this visual EDA tool, students inferred that the Nusselt number proportionally depends on heat capacity, diameter, density, and velocity, and is inversely proportional to the thermal conductivity and viscosity. If we recall the Reynolds and Prandtl numbers, the proportional variables identified through the correlogram are shown in the numerator. In contrast, the inversely proportional variables are shown in the denominator of these equations, proving the effectiveness of this tool while explaining the relationship between variables. Moreover, this correlogram allows for ranking the importance of these variables as we observe each circus radius (Pearson coefficient), with velocity, diameter, and thermal conductivity being the most influential variables, followed by viscosity, density, and heat capacity. Decision-making processes can be derived from this analysis when analysing process and equipment design or setting experiments.

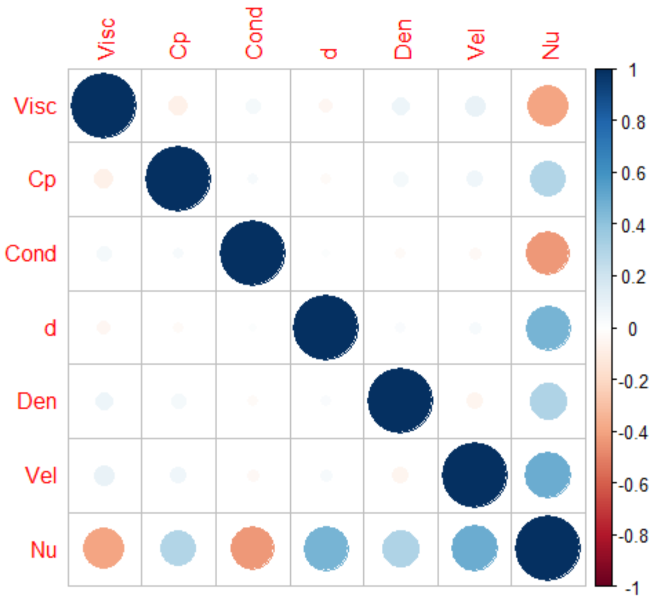

Figure 1: Correlogram for the relationship between variables affecting the Nusselt number.

Typical assessment questions during the lecture included: (i) describe the relationship between variables, (ii) describe the ranking of importance between the variables, and (iii) discuss the physical meaning and significance of the Nusselt number in heat transfer. While students were actively engaged in running the corresponding R code, we also noticed that the percentage of participation increased from 16% to 36% compared to two previous cohorts, measured as the number of participations per lecture. Some lessons-learned from this activity include (i) recognizing the importance of causation, visualized as the physical relationship between variables, and (ii) acknowledging causal relationships to support accurate outcome prediction.

A preliminary assessment of the implementation of this exercise shows an average improvement of +10% in the mark of the convection-related question in the final examination when discussing the problem results, compared with the previous cohort.

Impact of the Reaction Temperature on Conversion

Figure 2 shows the time series “conversion” and “temperature”, with an intervention effect (increasing the reaction temperature) that lifts the response variable (conversion) after 71 seconds.

To estimate a causal effect, the pre-intervention period is used for training the model, and the post-intervention period allows for computing a counterfactual prediction, showing how the response variable would have responded without the intervention.

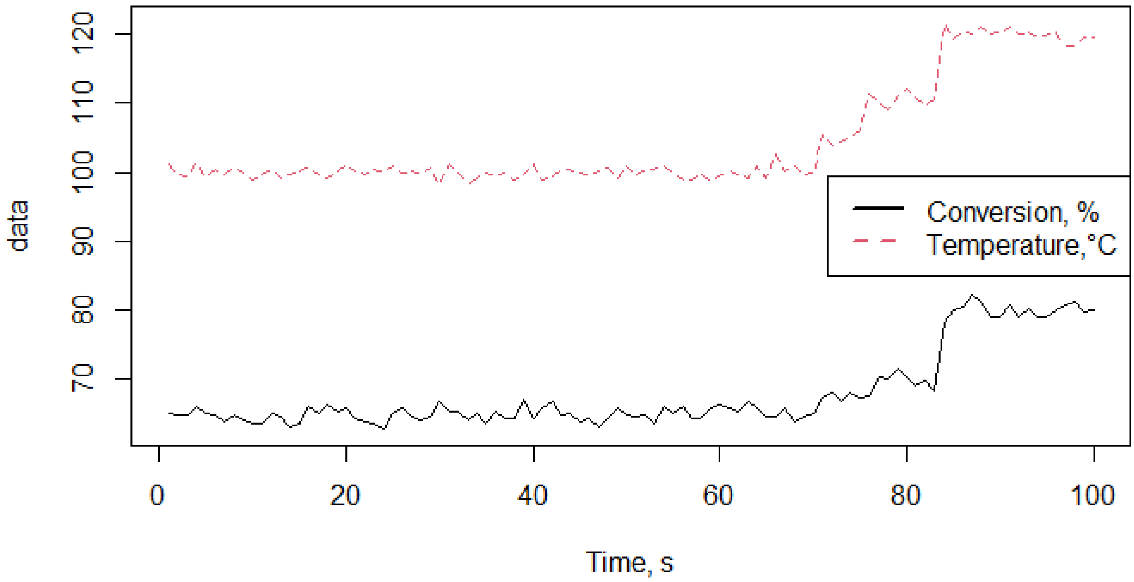

Figure 2: Reaction temperature and conversion time-series.

Figure 3 shows the impact of the intervention on the response in three panels: (i) the first panel shows the observed data and the predicted counterfactual during the post-treatment period; (ii) the second panel illustrates the pointwise causal effect, which is the difference between the observed data and the counterfactual predictions; (iii) the third panel shows the cumulative effect of the intervention, estimated as the adding up of the pointwise contributions. The counterfactual trend is clearly different than the one for the observed data, demonstrating the impact of the intervention on the response. This causation model assumes that the intervention does not impact covariates.

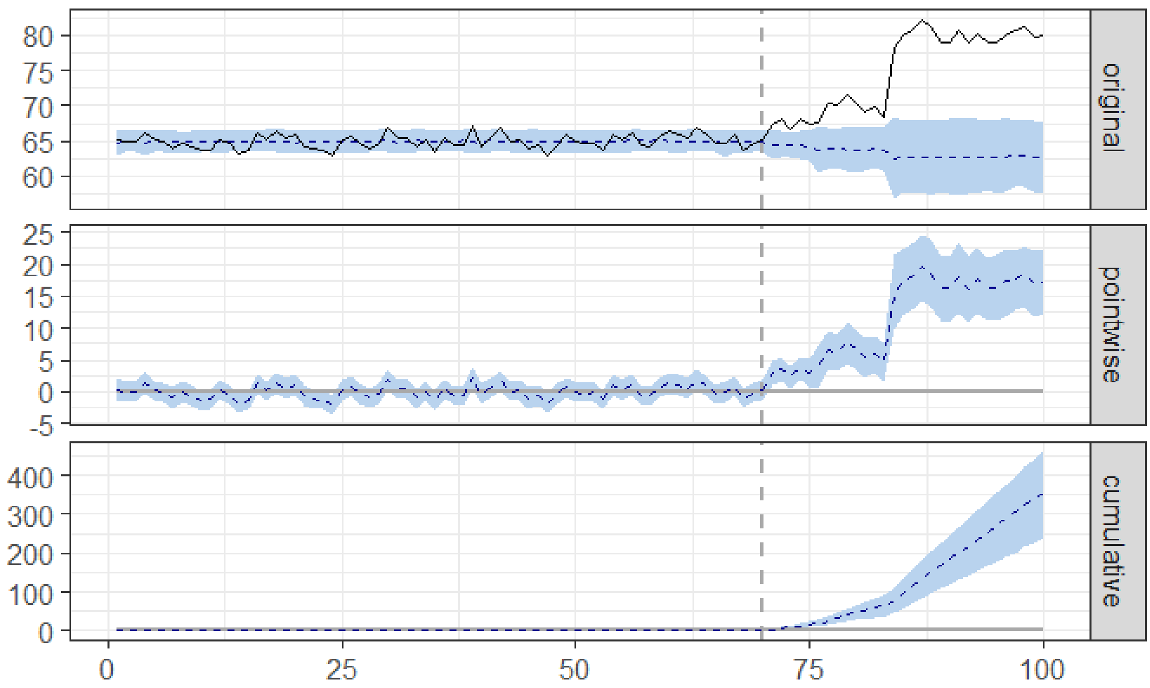

Figure 3: Impact of the intervention (temperature change).

Typical assessment questions during the lecture included: (i) describe the relationship between reaction temperature and conversion (based on the rate expression), (ii) describe the behaviour (trend) observed in the counterfactuals compared to the measured data, (iii) and the pointwise panel shows what appears to be a sub-intervention: should we analyse its effect as a standalone intervention? While students were actively engaged in running the corresponding R code, we also noticed that the percentage of participation increased from 20% to 40%, compared to two previous cohorts. Some lessons-learned gathered from this activity include (i) students recognized the importance of causation, visualized as an intervention in the time-series, (ii) data can provide valuable insights, but the underlying phenomenon must be understood, as engineers require informed decision-making, and (iii) potential use of this tool/method to investigate causes of unknown malfunctions to help expedite the problem-solving process.

A preliminary assessment of the implementation of this exercise shows an average improvement of +12% in the mark of the intervention-related question in the midterm examination when discussing the problem results, compared with the previous cohort.

Effect of the combined demographic variables and pollution exposure on health outcome

The demographic variables are inputted in our R code as categorical, while the treatment variable is continuous. The expected probability of the health outcome is set as binary.

The mean average of the ATE was estimated as 0.0171, greater than zero; the standard deviation (on 50 iterations) was 0.00035, proving the computational stability of the algorithm on randomly estimating the ATE. This value suggests that there is potential causation between demographic variables, exposure to a chemical pollutant, and the expected probability of the health outcome. Figure 4 shows a radar plot illustrating the dispersion of the ATE in 50 iterations.

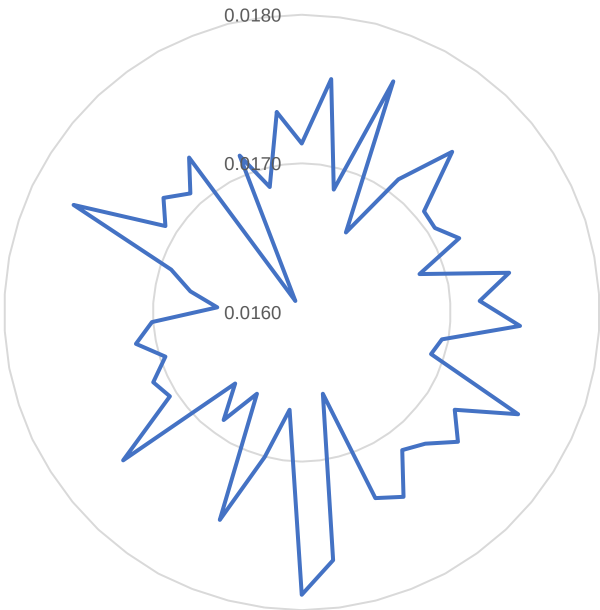

Figure 4: Radar plot showing the dispersion of the ATE in 50 iterations.

This example was assessed through a presentation, where the students described the problem statement, the machine learning method used for causation, results, and discussion, and overall feedback on the task. Some lessons-learned gathered among students include: (i) students recognized the importance of causation over a relationship, as correlation does not imply causation, (ii) there may be other factors (confounders) influencing variables, and (iii) causation is quite complex and difficult to prove as compared with a controlled experiment.

Future work

While the intended goals of applying our approach were attained, further works may consider evaluating the longitudinal effect and vertical implementation of this approach in our undergraduate chemical engineering curriculum, as several topics are either shared or pre-required in subsequent courses. Moreover, indicators around student retention are worthy of being researched while applying our approach, as well as the implicit applicability and impact of causation-based decision-making processes in more holistic and integrative courses such as the capstone-based ones.

Conclusions

This work presents a causation-driven approach to engineering education using data analytics and machine learning tools. With this approach, we aimed to increase student participation during lectures, encouraging students to recognize the importance of using data analytics and machine learning tools to understand phenomena and the cause-effect processes that can be understood through traditional teaching/learning approaches. The implementation of our approach was illustrated with three examples, including (i) understanding the physical significance of the Nusselt number using exploratory data analysis (through correlograms), (ii) studying the impact of the reaction temperature on conversion using Bayesian structural time-series model to evaluate the effect of an intervention, and (iii) using the Causal Random Forest algorithm to perform a heterogeneous treatment effect to assess the potential causation of combined demographic and environmental variables on health outcomes. The effectiveness of our approach was assessed by an increase of 20% in student participation compared with previous cohorts. Moreover, a set of lessons learned was gathered, attesting to insights that enhanced the student's learning experience. In the future, we intend to evaluate the longitudinal effect and vertical implementation of this approach in our undergraduate chemical engineering curriculum and conduct research around its potential impact on student retention, as well as the decision-making processes employed by students in integrative courses as capstone-based ones. Students' retention and quality of the learning/teaching process will also be assessed through surveys that compile general and subject-specific feedback evaluating aspects such as contribution to learning and motivation/awareness.

Acknowledgements

2023 Schwartz Reisman Institute for Technology and Society Faculty Fellowship.

Declaration of Interest

- All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revisiting it critically for impact intellectual content; and (c) approval of the final version.
- This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

- The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.
- The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Notes on Contributors

Daniela Galatro is an Assistant Professor, Teaching Stream, at the University of Toronto. Her research interests include engineering education, process simulation, and machine learning applied to process engineering.

Kai and Sathwik are undergraduate students in the Department of Chemical Engineering & Applied Chemistry at the University of Toronto. They enthusiastically joined the team of summer students, looking to deepen their technical skills on machine learning and proactively contribute to research.

References

Beebee, H., Hitchcock, C., Menzies, P., 2010. The Oxford Handbook of Causation. Oxford Academic.

Bermeo, C., Michell, K., Kristjanpoller, W., 2023. Estimation of causality in economic growth and expansionary policies using uplift modeling. Neural Comput Appl 35, 13631–13645. (https://doi.org/10.1007/s00521-023-08397-0)

Chakraborty, S., Gonzalez-Triana, Y., Mendoza, J., Galatro, D., 2023. Insights on mapping Industry 4.0 and Education 4.0. Front Educ (Lausanne) 8. (https://doi.org/10.3389/feduc.2023.1150190)

Colburn, A.P., 1964. A Method of Correlating Forced Convection Heat-Transfer Data and a Comparison with Fluid Friction. Pergamon Press.

Galatro, D., Chakraborty, S., Yan, N., Goodarzi, N., Castrucci, J., Saban, M. Education 4.0: Integrating Codes, Standards, and Regulations in the Chemical Engineering Curriculum. 2022: Proceedings of the Canadian Engineering Education Association (CEEA-ACEG) Conference June 18-22 York University. (https://doi.org/10.24908/pceea.vi.15829)

Kuo, C.L., Duan, Y., Grady, J., 2018. Unconditional or Conditional Logistic Regression Model for Age-Matched Case–Control Data? Front Public Health 6.(https://doi.org/10.3389/fpubh.2018.00057)

Liu, S., Glowatz, M., Zappatore, M., Gao, H., Jia, B., Bucciero, A., 2018. e-Learning, e-Education, and Online Training.

Machine Learning - Advanced Techniques and Emerging Applications, 2018. , Machine Learning - Advanced Techniques and Emerging Applications. InTech.

Prosperi, M., Guo, Y., Sperrin, M., Koopman, J.S., Min, J.S., He, X., Rich, S., Wang, M., Buchan, I.E., Bian, J., 2020. Causal inference and counterfactual prediction in machine learning for actionable healthcare. Nat Mach Intell 2, 369–375. (https://doi.org/10.1038/s42256-020-0197-y)

Russo, F., Wunsch, G., Mouchart, M., 2019. Causality in the Social Sciences: a structural modelling framework. Qual Quant 53, 2575–2588. (https://doi.org/10.1007/s11135-019-00872-y)

Sasmita, F., Mulyanti, B., 2020. Development of machine learning implementation in engineering education: A literature review. In: IOP Conference Series: Materials Science and Engineering. Institute of Physics Publishing.(https://doi.org/10.1088/1757-899X/830/3/032061)

Treviño-Elizondo, B.L., García-Reyes, H., 2022. What does Industry 4.0 mean to Industrial Engineering Education? In: Procedia Computer Science. Elsevier B.V., pp. 876–885. (https://doi.org/10.1016/j.procs.2022.12.284)

Zhao, Y., Liu, Q., 2023. Causal ML: Python package for causal inference machine learning. SoftwareX 21. (https://doi.org/10.1016/j.softx.2022.101294)