Real Data and Application-based Interactive Modules for Data Science Education in Engineering

Mr. Kerul Suthar, Auburn University

Kerul Suthar was born in Gujarat, India, in 1994. He received his B.S degree in chemical engineering from Dharmsinh Desai University in Gujarat, India in 2015. He is currently pursuing his Ph.D. degree in the department of chemical engineering at Auburn University, Auburn, Alabama. From 2015 to 2016, he was an operations engineer at Gujarat Narmada Valley Fertilizers and Chemicals Ltd in India. His research interests include the modeling and monitoring of manufacturing processes, Artificial Intelligence (AI), application of Machine learning (ML) techniques for smart and sustainable manufacturing, and Internet of Things (IoT) based sensing systems in the manufacturing industry.

Mr. Thomas Mitchell
Anna Claire Hartwig, Auburn University
Dr. Jin Wang, Auburn University

Dr. Jin Wang is B. Redd Associate Professor in the Department of Chemical Engineering at Auburn University. She obtained her BS and PhD degrees in chemical engineering (specialized in biochemical engineering) from Tsinghua University in 1994, and 1999 respectively. She then obtained a PhD degree (specialized in control engineering) from the University of Texas at Austin in 2004. From 2002 to 2006 she was a development engineer and senior development engineer at Advanced Micro Devices, Inc. During her tenure at AMD, her R&D yielded 13 patents granted by USPTO. In addition, she received several prestigious corporate awards for being instrumental in developing effective advanced control solutions. Dr. Wang joined Auburn University in 2006 as B. Redd Assistant Professor. She was promoted to Associate Professor and granted tenure in 2011. The central theme of her current research is to apply systems engineering, in particular, control engineering principles and techniques to understand, predict and control complex dynamic systems which cover both industrial processes and microbial organisms. Currently, she has extended her research focus to metabolic network modeling and analysis, as well as related experimentations. The system identification based framework for metabolic network analysis has been proving to be a highly effective tool to extract biological knowledge from complex, genome-scale metabolic network models, and has been successfully applied to understanding several industrial relevant microbes. She was the 2008 recipient of the Ralph E. Powe Junior Faculty Enhancement Awards from Oak Ridge Associated Universities (ORAU). Her graduate student also won the inaugural AIChE CAST Director's Presentation Award in 2011. Her research is funded by various US federal and state funding agencies including NSF, USDA, Department of Education and DOT as well as private foundations. She has over 40 journal publications, plus additional conference proceedings (>40) and presentations (>70). Her recent publications mainly focus on biotechnology and bioengineering related modeling and experimental research.

Prof. Shiwen Mao, Auburn University

Shiwen Mao received a Ph.D. in Electrical Engineering from Polytechnic University, Brooklyn, NY in 2004. He joined Auburn University, Auburn, AL in 2006 as an Assistant Professor in the Department of Electrical and Computer Engineering. He held the McWane Endowed Professorship from 2012 to 2015 and the Samuel Ginn Endowed Professorship from 2015 to 2020. Currently, he is a Professor and Earle C. Williams Eminent Scholar Chair, and Director of the Wireless Engineering Research and Education Center (WEREC) at Auburn University. Dr. Mao's research interest includes wireless networks, multimedia communications, and smart grid. He is a Distinguished Lecturer of IEEE Communications Society (2021-2022) and IEEE Council of RFID (2021-2022) , was a Distinguished Lecturer (2014-2018) and is a Distinguished Speaker (2018-2021) of IEEE Vehicular Technology Society. He received the IEEE ComSoc TC-CSR Distinguished Technical Achievement Award in 2019, the IEEE ComSoc MMTC Distinguished Service Award in 2019, the Auburn University Creative Research & Scholarship Award in 2018, the 2017 IEEE ComSoc ITC Outstanding Service Award, the 2015 IEEE ComSoc TC-CSR Distinguished Service Award, the 2013 IEEE ComSoc MMTC Outstanding Leadership Award, and the NSF

CAREER Award in 2010. He is a co-recipient of the 2021 IEEE Communications Society Outstanding Paper Award, the IEEE Vehicular Technology Society 2020 Jack Neubauer Memorial Award, the 2018 IEEE ComSoc MMTC Best Journal Paper Award, the 2017 IEEE MMTC Best Conference Paper Award, IEEE SECON 2017 Best Demo Award, Best Paper Awards from IEEE GLOBECOM 2019, IEEE GLOBECOM 2016, IEEE GLOBECOM 2015, IEEE WCNC 2015, and IEEE ICC 2013, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He is an Associate Editor-in-Chief of IEEE/CIC China Communications, an Area Editor of IEEE Transactions on Wireless Communications, IEEE Internet of Things Journal, IEEE Open Journal of the Communications Society, and ACM GetMobile, and an Associate Editor of IEEE Transactions on Cognitive Communications and Networking, IEEE Transactions on Network Science and Engineering, IEEE Transactions on Mobile Computing, IEEE Multimedia, IEEE Network, and IEEE Networking Letters. He is a Fellow of the IEEE, a member of the ACM, IET, Tau Beta Pi, and Eta Kappa Nu.

Laura Parson, North Dakota State University
Dr. Peng Zeng, Auburn University

Department of Mathematics and Statistics, Auburn University

Dr. Bo Liu
Dr. Peter He, Auburn University

Dr. Q. Peter He is Associate Professor in the Department of Chemical Engineering at Auburn University. He obtained his BS degree in chemical engineering from Tsinghua University, Beijing, China, in 1996 and MS and PhD degrees in chemical engineering in 2002 and 2005 from the University of Texas, Austin. Besides engineering education, his current research interests are in the area of systems engineering enhanced data analytics with applications in manufacturing, renewable energy, food-energy-water nexus, and broad area of disease detection/diagnosis and operation of healthcare systems. .

Real Data and Application based Interactive Modules for Data Science Education in Engineering

Kerul Suthar[1], Thomas Mitchell[1], Anna Hartwig[1], Jin Wang[1], Shiwen Mao[2], Laura Parson[3], Peng Zeng[4], Bo Liu[5], Q. Peter He[1,*]

[1]*Dept. of Chemical Engineering, Auburn University, Auburn, AL 36849*
[2]*Dept. of Electric and Computer Engineering, Auburn University, Auburn, AL 36849*
[3]*Dept. of Educational and Organizational Leadership, North Dakota State University, Fargo, ND 58105*
[4]*Dept. of Mathematics and Statistics, Auburn University, Auburn, AL 36849*
[5]*Dept. of Computer Science and Software Engineering, Auburn University, Auburn, AL 36849*
[*]*Corresponding author*

Abstract

The importance of data science and engineering (DSE) education cannot be overstated and undergraduate education offers a critical link in providing more DSE exposure to students and expanding the supply of DSE talent. Currently significant progress has been made in classwork, while progress in hands-on research experience is still lacking. To help fill this gap, we propose to create data-enabled engineering project (DEEP) modules in the form of interactive Jupyter Notebooks based on real data and applications. We hypothesize that this web-based interactive development and learning environment (IDLE) will enable easy and wide adoption of the DEEP modules by other educators and institutions. In this work, we will present our ideas, the rationale behind the proposed approach, the work in progress, and the future plans for the project.

Keywords

Data science, workforce development, data-enabled engineering project, experiential learning, course-based undergraduate research experience

Introduction

Data science is emerging as a field that is revolutionizing the world. A 2018 National Academies report – *Data Science for Undergraduates: Opportunities and Options* [1] states that "Work across nearly all domains is becoming more data driven, affecting both the jobs that are available and the skills that are required. As more data and ways of analyzing them become available, more aspects of the economy, society, and daily life will become dependent on data." In 2019, LinkedIn ranked "data scientist" the No. 1 most promising job in the U.S. based on job openings, salary, and career advancement opportunities and reported a 56% rise in job openings for data scientists over the previous year. The Bureau of Labor Statistics considers data science in the top 20 fastest growing occupations and has projected 31% growth between 2019-2029. Due to COVID-19, new job postings in data science and analytics have declined overall, however, they appear to be declining at a slower rate than that of most other occupations. And within the finance and insurance industry, new job postings in the analytics and data science space have actually increased.[2] The ever-increasing computing power, the exponential growth of data, and the desires of various industries and institutions to better use the data for informed decisions and optimal business outcomes, have been widely considered the reasons for the increasing demand of talents in data science and data analytics. To meet this increasing demand of data scientists and engineers, the National Academies report has recognized that undergraduate education offers

a critical link in providing more data science and engineering (DSE) exposure to students and expanding the supply of DSE talents. DSE education requires both appropriate classwork and hands-on experience with real data and real applications. While significant progress has been made in the former, one key aspect that has yet to be addressed is hands-on experience incorporating real-world applications. Specifically, it is insufficient for undergraduate students to be handed a "canned" data set and be told to analyze it using the methods that they are studying. Such an approach will not prepare them to solve more realistic and complex problems, especially those involving large, unstructured data. Instead, students need repeated practice with the entire DSE cycle beginning with ill-posed questions and "messy" data [1]. To this end, the following gaps have been identified. (1) There is a lack of real data and application based interactive learning materials for students to learn different aspects of DSE relevant to their life experiences and future job requirements. (2) There is lack of real data and application based research experiences or projects for students to practice the entire DSE workflow. To help fill these gaps, we have proposed to create interactive data-enabled engineering project (DEEP) modules based on real data and applications to be easily and widely adopted by other institutions. The rest of the paper is organized as follows. In Section 2, we briefly review some existing efforts and discuss their limitations. Section 3 presents the proposed approach. The work in progress is discussed in Section 4. The project future plans are presented in Section 5.

The existing efforts and their limitations

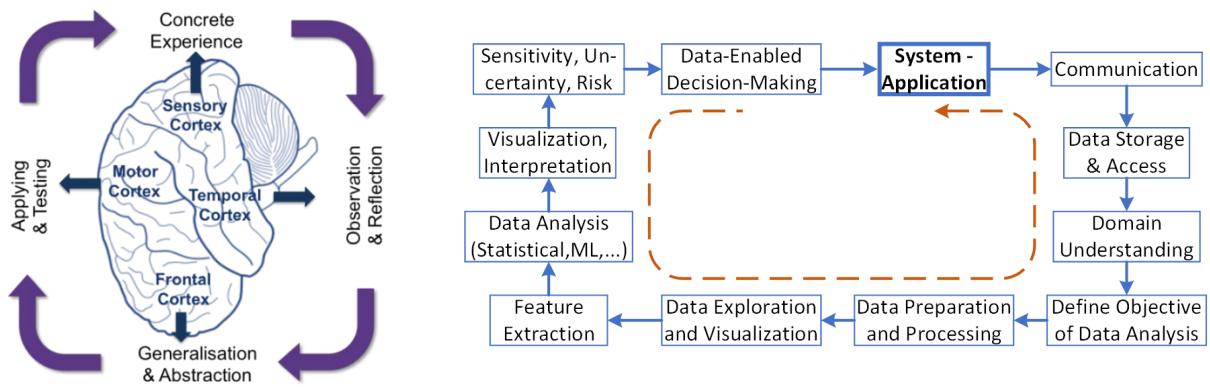
It has been recognized that textbooks and traditional lecture courses may offer limited help in developing students' capability in applying the theory and methods to solve real, complex problems [1]. There have been some efforts that integrate real projects and real data into DSE education. Some of them are listed in Table 1. However, there are limitations of these approaches including increased time, organizational and pedagogical demands, and other burdens on instructors[3], [4], challenges in solicitation of live projects [4], difficulty in finding assignments that motivate all students[5], possible lack of immediate applications, etc. More importantly, there is no learning material generated from these efforts that can be widely adopted for enhancing DSE education at other institutions.

Table 1. Some existing work on integrating real world data and applications into DSE education

Category	Representatives	Advantages	Limitations	
Live or real project based DSE course	Syracuse[6]: projects from instructor's clients; UT-Austin[7], Radford[4]: projects from on-campus centers; New Hampshire: projects from external partners[8]	Improve students' motivation and engagement; Promote teamwork; Improve students' written and oral communications skills	Increased time, organizational, and pedagogical demands, and other burdens on instructor[3], [4]; Solicitation of live projects is challenging, among others[4], [9], [10]	**No learning materials generated that can be widely adopted for enhancing DSE education at other institutions**.
Real data based DSE course	Smith College: Twitter[11]; Multiple institutions: DataFest[12] using multiple datasets of variety fields; Temple: multiple climate datasets[13]	Similar to those of live or real project based but to a lesser extent depending on the data source and domain expert support availability.	Difficult to find assignments that motivate all students[14]; May not have immediate applications	

The proposed approach

To address the above-mentioned limitations, we propose to develop data-enabled engineering project (DEEP) modules guided by the latest research on experiential learning theory (ELT). Experiential learning (EL) is the process of learning through experience, and is more specifically defined as "learning through reflection on doing"[15], [16]. Kolb helped to develop the modern theory of experiential learning, which focuses on the learning process of individual[15]. As shown in Fig. 1 (a), Kolb's EL is typically represented by a learning cycle: 1. Concrete Experience (*e.g.,* a new experience or situation); 2. Observation & Reflection of the new experience (*e.g.,* inconsistencies between experience and understanding); 3. Generalization & Abstraction (*e.g.*, reflection gives rise to a new idea, or modification of an existing one); 4. Applying & Testing (*e.g.*, apply idea(s) to the world). Fig. 1 (b) shows the entire DSE lifecycle, which shares the most important feature of EL cycle: it is a repeated cyclical process, and its different steps can be registered with different elements in EL cycle.



(a) Experiential learning cycle[17]

(b) DSE lifecycle

Fig. 1 (a) Experiential learning cycle; (b) a sample DSE lifecycle highlights the most important feature of experiential learning cycle: the learning of DSE lifecycle is an iterative cyclical and spiral process.

In addition, course-based undergraduate research experience (CURE) is a form of experiential learning that promotes all EL components in a positive cyclical and spiral learning process. As most DSE applications are open-ended research problems and learning an entire DSE lifecycle is really a research experience, the latest research on CURE provides excellent guidance for assembling DEEP modules into research projects. In particular, a 2019 study found that short CURE modules are an excellent alternative to more complex and costly whole-course CUREs and provide measurable metacognitive benefits to students[18]. Another benefit of short CURE modules is that they can be flexibly inserted into existing curricula[18], [19]. Therefore, we further propose to adapt the short CURE module mechanism to assemble DEEP modules into short DEEP-CUREs.

Work in progress

Sources of real data and applications

To achieve the project goals, we have developed two internet-of-things (IoT) enabled laboratory engineering testbeds and generated real data under various application scenarios[20]–[22]. Another important source of real data and applications is the sample datasets from MathWorks, which are listed in Table 2. As can be seen, most of these datasets have science and engineering background and applications, which make them especially suitable for this project. In addition, there are publicly available datasets, including those from Kaggle (https://www.kaggle.com/datasets), Amazon (https://registry.opendata.aws/), Google (https://cloud.google.com/bigquery/public-data/), the U.S. government (https://www.data.gov/), and many others. Some of them may be suitable for DEEP-CUREs, but we have not looked into them carefully.

Table 2. MathWorks sample datasets (https://www.mathworks.com/help/stats/sample-data-sets.html)

File	**Description of Dataset**
acetylene.mat	Chemical reaction data with correlated predictors
arrhythmia.mat	Cardiac arrhythmia data from the UCI machine learning repository
carbig.mat	Measurements of cars, 1970–1982
carsmall.mat	Subset of carbig.mat. Measurements of cars, 1970, 1976, 1982
census1994.mat	Adult data from the UCI machine learning repository
cereal.mat	Breakfast cereal ingredients
cities.mat	Quality of life ratings for U.S. metropolitan areas
discrim.mat	A version of cities.mat used for discriminant analysis
examgrades.mat	Exam grades on a scale of 0–100
fisheriris.mat	Fisher's 1936 iris data
flu.mat	Google Flu Trends estimated ILI (influenza-like illness) percentage for various regions of the US, and CDC weighted ILI percentage based on sentinel provider reports
gas.mat	Gasoline prices around the state of Massachusetts in 1993
hald.mat	Heat of cement vs. mix of ingredients
hogg.mat	Bacteria counts in different shipments of milk
hospital.mat	Simulated hospital data
humanactivity.mat	Human activity recognition data of five activities: sitting, standing, walking, running, and dancing
imports-85.mat	1985 Auto Imports Database from the UCI repository
ionosphere.mat	Ionosphere dataset from the UCI machine learning repository
kmeansdata.mat	Four-dimensional clustered data
lawdata.mat	Grade point average and LSAT scores from 15 law schools
mileage.mat	Mileage data for three car models from two factories
moore.mat	Biochemical oxygen demand on five predictors
morse.mat	Recognition of Morse code distinctions by non-coders
nlpdata.mat	Natural language processing data extracted from the MathWorks® documentation
ovariancancer.mat	Grouped observations on 4000 predictors [1][2]
parts.mat	Dimensional run-out on 36 circular parts
polydata.mat	Sample data for polynomial fitting
popcorn.mat	Popcorn yield by popper type and brand
reaction.mat	Reaction kinetics for Hougen-Watson model
spectra.mat	NIR spectra and octane numbers of 60 gasoline samples
stockreturns.mat	Simulated stock returns

Implementation approaches

One major obstacle of integrating DSE education into all STEM disciplines is that not all academic programs will be able to accommodate the addition of a designated DSE course with all other programmatic requirements currently in place. Another obstacle is that any changes to the curriculum that require significant effort from faculty or staff would be difficult to sell and would not be adopted widely. In the proposed framework, the DEEP modules serve as the supplementary materials to existing STEM courses. In other words, the basic concepts and fundamental principles to be learned remain the same and are taught in the same way as in a traditional STEM class. The DEEP modules can be used to replace/supplement some of the textbook examples and homework problems. Therefore, the proposed DEEP module approach can effectively address the above-mentioned obstacles.

Since its introduction, Jupyter Notebook has become an excellent interactive training platform for various fields, especially those involving computing, data science and machine learning [23]–[27]. The main advantage of Jupyter Notebook is that it allows the combination of markup language, graphics, and equations with interactive, executable Python codes (and other programming languages such as R, SAS, and Julia). In addition, a Jupyter Notebook can be converted to a number of open standard output formats (HTML, presentation slides, LaTeX, PDF, ReStructuredText, Markdown, Python) through "Download As" in the web interface, which is convenient to generate homework submissions or project reports. These characteristics make Jupyter Notebook a powerful and easy-to-grasp tool for students to develop entire projects, and instructors and trainers to develop efficient and interactive training materials. Finally, the Jupyter Notebook has become a popular user interface for cloud computing. As a result, major cloud providers have adopted the Jupyter Notebook as a frontend interface for cloud users. Examples include Amazon's SageMaker Notebooks, Google's Colaboratory (the platform selected by this project), and Microsoft's Azure Notebook. Jupyter Notebooks hosted on the platforms mentioned above can be shared via a link, accessed and executed on any web browser of any operation systems without any downloads or local software/package installations. This is important as local software/package installations can present significant technical and psychological barriers for the adoption of the DEEP modules by the instructors and institutions, as well as the utilization level of the students.

Sample modules developed

We have completed three modules under the subject of Data Preparation and Processing: (1) Missing Value; (2) Noise; and (3) Outlier. Within each module, there are six sections, including overview, detection, handling, impact on machine learning, problems and references. We have also developed modules on Data Visualization, Clustering, and Classification. One sample Jupyter Notebook screenshot is shown in Fig. 2.

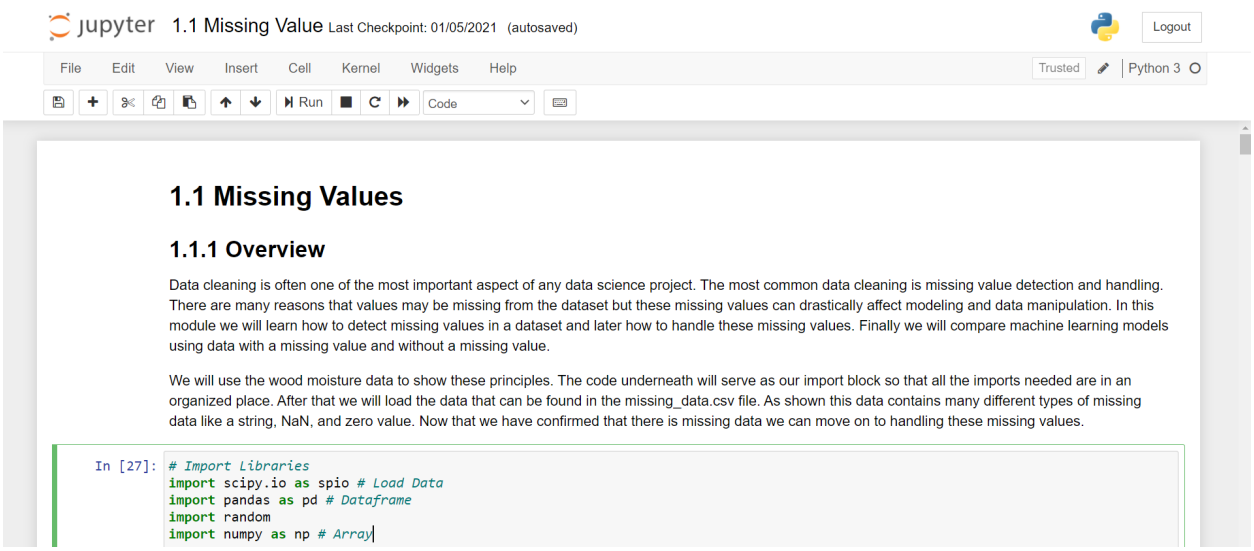


Fig. 2 Screenshot of a sample DEEP module in Jupyter Notebook: missing values, their detection and handling, and their impact on some common machine learning tasks

As different courses/instructors may use different programming languages, we have ported the developed DEEP modules to MATLAB using MATLAB Live Scripts, which are interactive documents that combine MATLAB code with formatted text, equations, and images in a single environment, very much similar to the Jupyter Notebook except that a MATLAB license is required. One sample MATLAB Live Script screenshot is shown in Fig. 3.

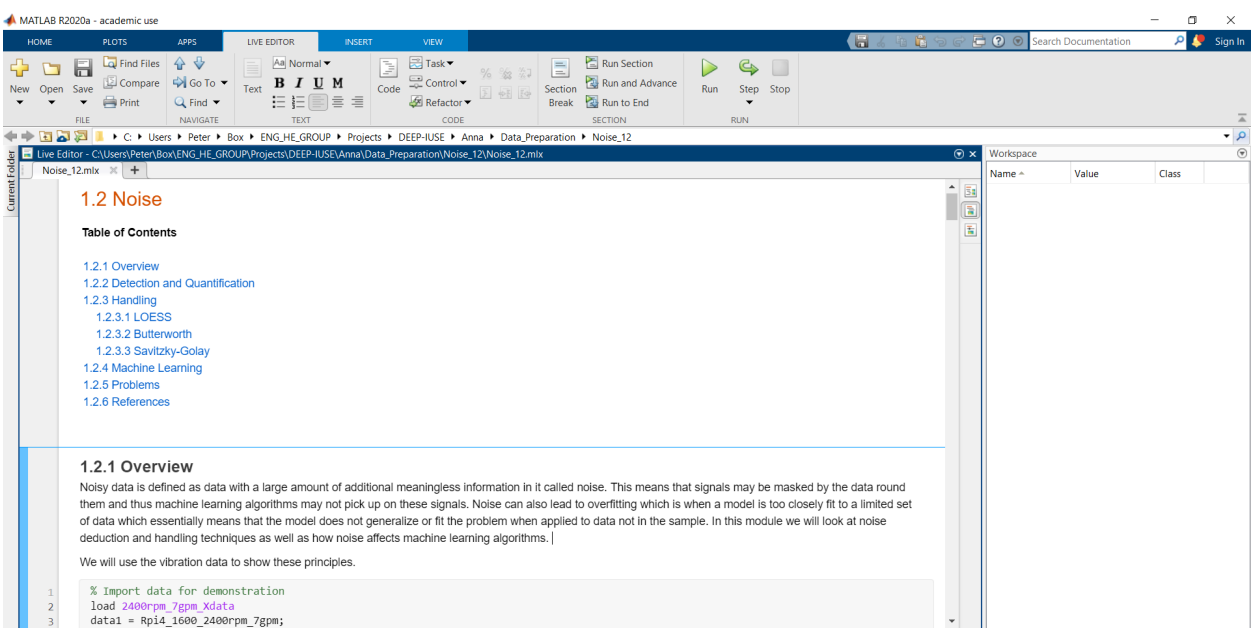


Fig. 3 Screenshot of a sample DEEP module in MATLAB Live Script.

Some DEEP modules have also been developed in SAS via the SASPy Python package. A screenshot of one sample DEEP module in SAS is shown in Fig. 4. The link “Open in Colab”

shown in the upper part of Fig. 4 allows single click to open it in Google Colab for interactive execution and testing.

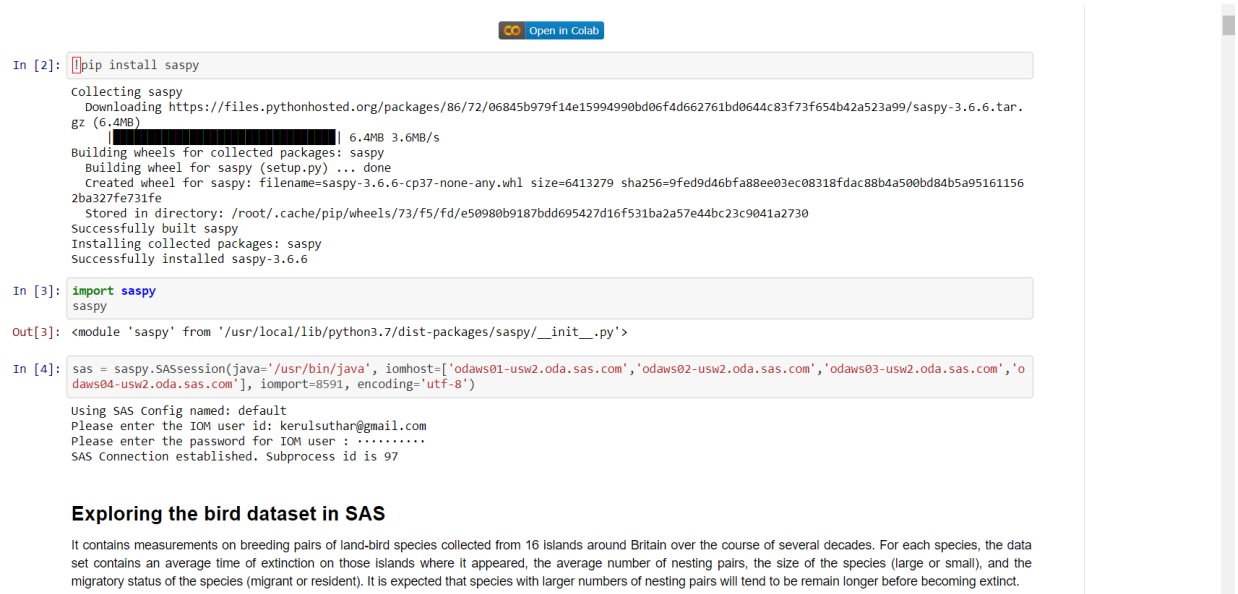


Fig. 4 Screenshot of a sample DEEP module in SAS.

At this point, all modules are kept private on GitHub as they are still in the development phase. Once they are tested and finalized with comments and feedback addressed, they will be made public and advertised via different venues (*e.g.*, ASEE publication/workshop) to reach other educators for broad dissemination.

Future plans

We plan to complete DEEP and DEEP-CURE module development by the summer of 2021, then test them in six engineering courses from four different disciplines: Chemical Engineering, Electrical and Computer Engineering, Computer Sciences, and Mathematics and Statistics at Auburn University. We hypothesize that ELT guided DEEP modules, especially those designed for students to answer important reflective questions in DSE, together with DEEP-CUREs, will significantly enhance students' reflection and metacognition. Metacognition refers to the ability to reflect upon, understand, and control students' learning[28], [29]. Extensive research has indicated that metacognitively aware learners are more strategic and perform better than unaware learners[28]–[34]. In this project, a widely used 52-question Metacognition Awareness Inventory (MAI) will be used to quantify students' metacognition awareness gains. The scores will be analyzed and compared categorically and holistically to test our above-mentioned hypothesis. By reviewing the MAI questions, we expect that DEEPs will mostly enhance the "regulation" aspect of metacognition. However, it is possible that the "knowledge" aspect, such as knowledge about strategies and how to use strategies could be enhanced as well. The findings will add evidence to CURE's effectiveness in DSE education and in enhancing metacognition in general. The findings will also add knowledge to ELT in DSE education, and provide a model of ELT guided DEEP module design for addressing specific aspects of ELT.

References

[1] E. National Academies of Sciences and Medicine, *Data science for undergraduates: opportunities and options*. National Academies Press, 2018.

[2] J. Camm, M. Bowers, and T. Thomas, "The Recession's Impact on Analytics and Data Science," *MIT Sloan Management Review*, p. 1, 2020.

[3] D. D. Bickerstaff Jr, "The evolution of a project oriented course in software development," *ACM SIGCSE Bull.*, vol. 17, no. 1, pp. 13–22, 1985.

[4] J. D. Chase, E. Oakes, and S. Ramsey, "Using live projects without pain: the development of the small project support center at Radford University," *ACM SIGCSE Bull.*, vol. 39, no. 1, pp. 469–473, 2007.

[5] C. J. Ballen *et al.*, "A call to develop course-based undergraduate research experiences (CUREs) for nonmajors courses." Am Soc Cell Biol, 2017.

[6] J. Saltz and R. Heckman, "Big Data science education: A case study of a project-focused introductory course," *Themes Sci. Technol. Educ.*, vol. 8, no. 2, pp. 85–94, 2016.

[7] P. S. Grisham, H. Krasner, and D. E. Perry, "Data engineering education with real-world projects," *ACM SIGCSE Bull.*, vol. 38, no. 2, pp. 64–68, 2006.

[8] M. Sabin, "A collaborative and experiential learning model powered by real-world projects," in *Proceedings of the 9th ACM SIGITE conference on Information technology education*, 2008, pp. 157–164.

[9] D. Farkas, "Choosing group projects for advanced systems courses," in *ACM SIGCSE Bulletin*, 1988, vol. 20, no. 1, pp. 109–115.

[10] W. J. Joel, "Realistic student projects," *ACM SIGCSE Bull.*, vol. 19, no. 1, pp. 244–247, 1987.

[11] B. Baumer, "A data science course for undergraduates: Thinking with data," *Am. Stat.*, vol. 69, no. 4, pp. 334–342, 2015.

[12] R. Gould, "Datafest: Celebrating data in the data deluge," 2014.

[13] G. Witt, "Using data from climate science to teach introductory statistics," *J. Stat. Educ.*, vol. 21, no. 1, 2013.

[14] J. Hardin *et al.*, "Data science in statistics curricula: Preparing students to 'think with data,'" *Am. Stat.*, vol. 69, no. 4, pp. 343–353, 2015.

[15] D. Kolb, "Towards an applied theory of experiential learning," *Theor. Gr. Process.*, pp. 33–56, 1975.

[16] P. Stonehouse, P. Allison, and D. Carr, "Aristotle, Plato, and Socrates: Ancient Greek perspectives on experiential learning," *Sourceb. Exp. Educ. Key thinkers their Contrib.*, pp. 18–25, 2011.

[17] J. E. Zull, *The art of changing the brain: Enriching teaching by exploring the biology of learning*. Stylus Publishing, LLC., 2002.

[18] C. L. Dahlberg *et al.*, "A Short, Course-Based Research Module Provides Metacognitive Benefits in the Form of More Sophisticated Problem Solving.," *J. Coll. Sci. Teach.*, vol. 48, no. 4, 2019.

[19] D. R. Howard and J. A. Miskowski, "Using a module-based laboratory to incorporate

inquiry into a large cell biology course," *Cell Biol. Educ.*, vol. 4, no. 3, pp. 249–260, 2005.

[20] K. Suthar, J. Wang, Z. Jiang, and Q. P. He, "Using Channel State Information for Estimating Moisture Content in Woodchips via 5 GHz Wi-Fi," in *2020 American Control Conference (ACC)*, 2020, pp. 2784–2789.

[21] D. Shah, J. Wang, and Q. P. He, "An Internet-of-things Enabled Smart Manufacturing Testbed," in *IFAC-PapersOnLine*, 2019, vol. 52, no. 1, pp. 562–567, doi: 10.1016/j.ifacol.2019.06.122.

[22] D. Shah, J. Wang, and Q. P. He, "Knowledge-Guided Feature Engineering in Big Data Analytics for IoT-Enabled Smart Manufacturing," *Comput. Chem. Eng.*, p. Under review, 2019.

[23] S. Farrell *et al.*, "Interactive Distributed Deep Learning with Jupyter Notebooks," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11203 LNCS, pp. 678–687, 2018, doi: 10.1007/978-3-030-02465-9_49.

[24] H. Igel, B. Chow, S. Donner, L. Krischer, M. van Driel, and C. Tape, "Jupyter Notebooks for Earth Sciences: An Interactive Training Platform for Seismology," 2017.

[25] A. Sanchez Pineda and A. R. Sanchez Pineda, "ATLAS OpenData and OpenKey: using low tech computational tools for students training in High Energy Physics," ATL-COM-SOFT-2018-053, 2018.

[26] A. Suárez, M. A. Alvarez-Feijoo, R. Fernández González, and E. Arce, "Teaching optimization of manufacturing problems via code components of a Jupyter Notebook," *Comput. Appl. Eng. Educ.*, vol. 26, no. 5, pp. 1102–1110, 2018.

[27] F. Perez and B. E. Granger, "Project Jupyter: Computational narratives as the engine of collaborative data science," *Retrieved Sept.*, vol. 11, no. 207, p. 108, 2015.

[28] G. Schraw and R. S. Dennison, "Assessing metacognitive awareness," *Contemp. Educ. Psychol.*, vol. 19, no. 4, pp. 460–475, 1994.

[29] G. Schraw, "Promoting general metacognitive awareness," *Instr. Sci.*, vol. 26, no. 1–2, pp. 113–125, 1998.

[30] A. Young and J. D. Fry, "Metacognitive awareness and academic achievement in college students.," *J. Scholarsh. Teach. Learn.*, vol. 8, no. 2, pp. 1–10, 2008.

[31] S. Turan, Ö. Demirel, and İ. Sayek, "Metacognitive awareness and self-regulated learning skills of medical students in different medical curricula," *Med. Teach.*, vol. 31, no. 10, pp. e477–e483, 2009.

[32] F. ur Rahman, N. B. Jumani, M. A. Chaudry, and F. Abbasi, "Impact of metacognitive awareness on performance of students in chemistry," *Contemp. Issues Educ. Res.*, pp. 39–44, 2010.

[33] É. Kállay, Learning strategies and metacognitive awareness as predictors of academic achievement in a sample of romanian second-year students., *Cogn. Creier, Comport. Brain, Behav.*, vol. 16, no. 3, 2012.

[34] E. Yeşilyurt, "Metacognitive Awareness and Achievement Focused Motivation as the Predictor of the Study Process.," *Int. J. Soc. Sci. Educ.*, vol. 3, no. 4, 2013.