DaST: An Online Platform for Automated Exercise Generation and Solving in the Data Science Domain

Charis Kotsiopoulos
kotsiopoulos@uop.gr
University of the Peloponnese, Tripoli, Greece

Ioannis Doudoumis
dit15035@uop.gr
University of the Peloponnese, Tripoli, Greece

Paraskevi Raftopoulou
praftop@uop.gr
University of the Peloponnese, Tripoli, Greece

Christos Tryfonopoulos
trifon@uop.gr
University of the Peloponnese, Tripoli, Greece

ABSTRACT

Over the last few years data science has emerged both as a new research field and as an educational domain that attracted a large number of researchers and data practitioners. Although data science research is developing at a high pace, the educational process in the field has been left behind in terms of educational tools and practices, despite the high number of data science courses offered and the number of involved stakeholders (professors, tutors, and students). The present work aims to cover the gap of educational data science tools by proposing a novel platform; the platform, coined Data Science Tutor (DaST), is a free online tool that offers automated step-by-step exercise solving in a variety of data science algorithms/techniques aiming at giving insight to the particularities of each algorithm. The solutions of the exercises are accompanied with in-context explanations that refer to the operation of the respective algorithm/technique, and are compatible with the terminology and the methodology in popular textbooks. The tool aims at students, lecturers, and data practitioners in many diverse fields (ranging from data analysts to transport engineers to logistics managers) that want to learn the particularities of data science algorithms in a stepwise, interactive manner.

Through the proposed platform (a) students in the data science field or in related courses (e.g., machine learning, information retrieval) may get solutions for different types of exercises and focus on the details of each algorithm, (b) tutors (lecturers, lab assistants) may easily produce a wide variety of exercises with the accompanying solutions and use them in classroom or as an auxiliary tool for test correction, and (c) data practitioners may get valuable insights on popular data science algorithms. To the best of our knowledge, the proposed platform is the first educational tool that aims at the data science field, and has so far been warmly accepted by departments worldwide.

CSERC '19, November 18–20, 2019, Larnaca, Cyprus

ACM ISBN 978-1-4503-7717-1/19/11…$15.00
https://doi.org/10.1145/3375258.3375271

CCS CONCEPTS

• **Information systems** → *Data management systems*; • **Applied computing** → **Education**; **Computer-assisted instruction**; **Interactive learning environments**; **Collaborative learning**.

KEYWORDS

data science, online platform, exercise solving, tertiary education

ACM Reference Format:
Charis Kotsiopoulos, Ioannis Doudoumis, Paraskevi Raftopoulou, and Christos Tryfonopoulos. 2019. DaST: An Online Platform for Automated Exercise Generation and Solving in the Data Science Domain. In *Proceedings CSERC 2019 18-20 November 2019 Computer Science Education Research Conference Larnaca, Cyprus (CSERC '19), November 18–20, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3375258.3375271

1 INTRODUCTION

Lately, the explosion of available data and the need to extract insight out of this data triggered the creation of a new interdisciplinary field that involved areas ranging from applied mathematics to statistics and from artificial intelligence to machine learning. This field, known as data science [5, 6, 15, 25], stimulated changes both in research and education in universities around the world. Towards this new direction, basic and applied research has focused on ways of managing, analysing, extracting, and visualising the information that resides behind the data, and utilising it in decision-making.

Naturally, data science requires skilled and appropriately trained data scientists, and this need led several departments to incorporate data science related courses, e.g., big data management, in their (undergraduate or postgraduate) studies program. These courses were offered alongside traditional information/data management courses such as machine learning or information retrieval, thus creating a large educational "market" involving a significant number of people (tutors, lab assistants, students) that teach or study methods, algorithms, and techniques in the broad area of data science [4, 11]. Despite this explosion in interest for data science related studies, there is a severe lack of educational tools that could assist the process of teaching and learning in such a broad and rapidly evolving scientific field [8, 11, 22, 26].

This paper presents a *novel online platform*, coined Data Science Tutor (DaST), that is used for automated problem solving in the data science domain. Users may automatically design and solve exercises for data science algorithms by resorting to a completely free and easy-to-use tool. The solutions of the exercises are presented in a step-by-step manner, followed by explanations that refer to

the operation of the respective algorithm or technique, and are compatible with the terminology and the methodology in popular textbooks, like [1, 3, 17, 19]. The DaST platform acts as a computer-assisted tutor for a wide range of data management techniques and algorithms related to the broader field of data science, where trainers and trainees may create their own exercises, resort to the ready-to-use example ones, or solve exercises submitted by other users. The proposed platform targets two main categories of users, revealing the different objectives for each category. Undergraduate or postgraduate students in the field of data science (i) have the potential to learn the particularities of each algorithm and identify cases that are not covered by the in-class teaching material and (ii) study in an interactively assisted way the reasoning behind problem solving and the exercise solution process for different types of exercises. Similarly, through the proposed platform, the tutors of a course (lecturers or lab staff) may easily produce a wide variety of exercises, with the accompanying solutions, and use them in the classroom or as an auxiliary tool for producing and correcting student tests. The DaST platform is in use for about one year and is continuously enriched with new functionality and algorithms. To the best of our knowledge, the DaST platform is the *only online educational tool* available in the data science domain and has already been warmly accepted by faculties worldwide.

The rest of the paper is organised as follows. In Section 2, we present the available approaches related to our platform. Section 3 describes the architecture, the capabilities and the provided functionality of the DaST platform, while in Section 4, we present and discuss the user acceptance of the proposed platform. Finally, Section 5 gives directions for future research and concludes the paper.

2 RELATED WORK

Our research is broadly related to approaches that deal with the organization and management of the study, including educational portals and online learning environments. In [20] an educational portal based on personalized information retrieval is presented; the proposed solution aims to act as an aid at the learning process by retrieving information pertaining to the context of the problems studied by the students. In a similar spirit, [27] visualizes a user model to achieve adaptive information retrieval in a learning environment, while [12] addresses effective content-delivery with the help of a university intranet, and collates the teacher-moderated online forum with the associated intranet portal. Online educational tools for guiding learners through the programming process have also been lately proposed; [14] focuses on the programming process by providing a tool that helps students learn the necessary steps towards designing a computer program, while [21] concentrates on JavaScript. Finally, [7] presents the educational goals and objectives when dealing with a broad teaching domain and aims at student collaboration for analyzing and tackling difficult problems.

Works that are most closely connected to our platform include [9, 10, 16]. In [9] an educational system that assists students in learning and tutors in teaching is presented; however, this system deals only with search algorithms within the artificial intelligence (AI) domain. Towards a different direction, [10] proposes a computer-supported learning environment using an information retrieval (IR) game, aiming to provide a realistic environment for demonstrating the performance of queries in different types of search situations. The experimental evaluation of this system revealed that students found different characteristics of the IR game both enhancing and inhibiting learning. Similarly, [16] presents an AI game platform, which includes state-of-the-art algorithms and aims to bring a rich and fun learning experience in the AI domain. Finally, problem meta-heuristic solver [13] is an educational software tool that aims at the generic study of the concepts related to the optimization field, by covering the main stages when solving optimization problems; the results obtained by its usage suggest that the tool improves the understanding of the theoretical concepts and reduces the workload of the students when implementing optimization methods.

Educational tools prove important not only to the computer science field, but also to other fields like mathematics or art. Wolfram Alpha is a platform similar to DaST, that targets exercise solving in the field of mathematics [2, 18, 23]. This platform is a computational knowledge engine that is able to answer factual mathematical queries, providing (with a subscription fee) a step-by-step solution for a wide variety of mathematical problems. Maestoso [24] is an intelligent computer-assisted educational application that allows people to learn the theory of music through sketching, and enables them to progress through provided lessons of important fundamentals in music theory.

Although all the above approaches concern different aspects of on-line student aids, none of them focuses on computer-assisted learning on the data science field, despite the fact that research [22] has indicated the need for such tools in a rapidly evolving domain that involves a broad spectrum of techniques and algorithms. This need for learning environments related to data science inspired us to design and develop the DaST platform as an *on-line, free-of-charge tool* that will enhance the students' learning process and assist tutors in their teaching tasks.

3 THE DAST PLATFORM

In this section, we describe the architecture, the capabilities and the quality of the provided functionality for the DaST platform.

The DaST platform has been implemented using widespread programming languages and tools, while its architecture is stratified with the goal of creating an easy way to add new techniques and algorithms. It is entirely based on php and javascript, fully compatible with the latest HTML and CSS standards, while for displaying widgets (e.g., charts or graphs) it uses the Google Chart API. At this stage, the platform is hosted on a typical Ubuntu Linux (version 16.04.3) server with 64bit / Dual Core 2GHz / 4GB RAM.

3.1 Features and functionality

Access to the platform is possible via any modern browser and is provided for free, does not require any sort of registration, and does not collect any type of personal information. Thus, by just typing the web address of the DaST platform[1] the user is transferred to the graphical interface of the proposed application (Figure 1) and may benefit from all the functionality provided by the DaST platform. There, a variety of algorithms and techniques related to the data science domain are presented, each alongside three different choices meant for creating, and subsequently automatically solving, an exercise. The user may choose the algorithm/technique he is

[1] http://db.uop.gr/~dast/

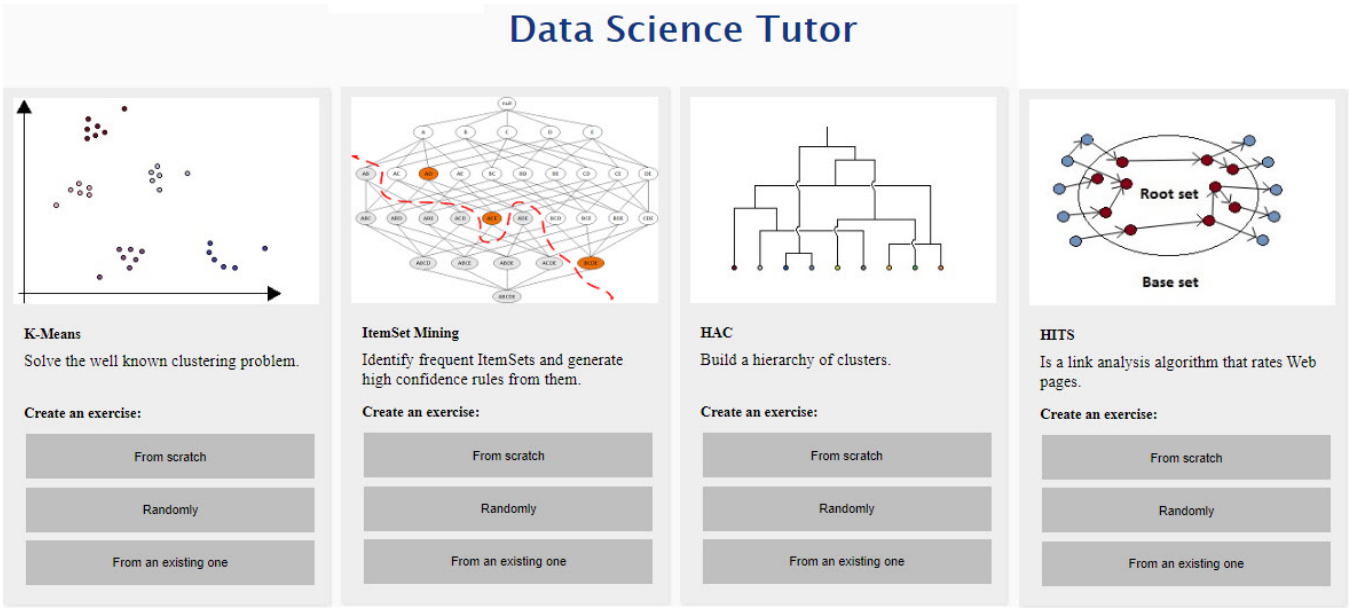


Figure 1: Part of the home screen of the DaST platform presenting (some of) the available algorithms

interested in practicing or for producing auxiliary material, and may also select to create the exercise from scratch, get a randomly generated one, or by resorting to an already submitted exercise by another user. In its current version, the DaST platform provides the opportunity to produce, practice with, and solve several types of exercises for fourteen (14) different algorithms alongside their variants. The DaST platform provides the users with the choice to print or export to a PDF file the exercise with the provided step-by-step solution. The whole interaction with the platform, as well as the provided step-by-step solutions and comments of the exercises are in English, and all the provided information on the algorithms references the relevant literature.

All algorithms and techniques available in the DaST platform are accompanied with a concise presentation of the related theory as well as assistance for the type of entry required to create an exercise for each algorithm or technique. It is worth mentioning that the DaST platform addresses the tertiary education and focuses on a broad and demanding field of science; the algorithms and techniques related to the data science domain require specialized knowledge, even for the data entry needed to create an exercise. For this reason, the provision of a guide for each algorithm/technique was deemed necessary after the requirements analysis and personal interviews conducted to users of the platform. The guide provided by the DaST platform for each algorithm or technique contains all the necessary explanations that refer to its operation (and to the step-by-step solutions of the corresponding exercises), is fully compatible with the terminology and the methodology met in popular textbooks, and is used within a wizard-style environment to facilitate the exercise creation process.

In what follows, we present the three different ways the user may utilise to create an exercise, after having chosen an algorithm or technique provided by the DaST platform.

3.1.1 Exercise creation from scratch. In this case, the user may create an exercise by providing/uploading at the DaST platform his own data. The process for entering the necessary input data for each algorithm is done using a simple, user-friendly wizard that guides the user in filling in the required fields. User data input is supported in three ways: (i) by using the online graphical environment, (ii) by uploading an appropriately formatted text file containing the input data (parameters still have to be inserted through the graphical interface), or (iii) by uploading an appropriately formatted XML file containing all the necessary parameter setups and data input that will allow the creation of the exercise (and its subsequent solution). The different options for providing input address the needs of both beginner/occasional users and also more expert ones who often require an easy, fast, and semi-automated way to create exercises. For all file-based input the system provides example (.txt and .xml) files for downloading, while to support interoperability with other applications the XML schema is also available through the platform.

3.1.2 Random exercise creation. In this case, the chosen algorithm or technique is fed with randomly generated data and parameters. Choosing this way to create an exercise, the interested user has the opportunity to understand the functionality of an algorithm by studying the step-by-step solution of an exercise, without being puzzled with the data entry to create the exercise. This choice mostly targets students or beginners that are looking for a simple and straightforward way to familiarise with data science concepts and methods through exercise solving. For repeatability reasons the random exercise generation process asks the user to input a seed that may be used again later on to recreate the exercise. If none is provided, the system generates one seed by resorting to a time variable, and reports it to the user for future reference.

3.1.3 Exercise creation based on existing ones. In this case, the user may select to use and solve one of the exercises that were previously submitted by another user. Please notice that user anonymity is fully

protected as users are not required to register to the platform and the application does not relate in any way which user created which exercise. The users may change the data and/or the parameters of submitted exercises and study the way the algorithm and the provided solution are affected. The selection between the existing exercises is performed either randomly or based on a specific feature (e.g., difficulty of the exercise, type of user –tutor or student– that created it). Using exercises provided by others, while protecting the anonymity of the users, provides the DaST platform with an interesting social dimension. This feature is currently under testing.

3.2 Supported algorithms

The full list of algorithms that are available in the current edition of the DaST platform are presented below. The algorithms are classified here in research fields related to the data science domain.

3.2.1 Unsupervised clustering. Unsupervised clustering identifies previously unknown patterns in data sets without pre-existing labels. In essence, unsupervised clustering involves a set of simple-yet-powerful tools for identifying groups of objects, and typically serves as an introduction for more complicated techniques that involve also (semi-)supervised counterparts.

K-Means is the most well-known algorithm for unsupervised clustering of objects; objects in DaST are represented as vectors with up to 10 dimensions – remember that the aim is exercise solving, not real-world operation. Users input the vectors and the algorithm setup. The number of dimensions is automatically inferred (vectors with less dimensions are padded with zeros), while the user may tune the number of clusters, the number of iterations, the initial cluster seeds, the distance method (Euclidean or Manhattan), and the algorithm variant (sequential or on-line).

K-Medoids is another well-known unsupervised clustering algorithm that is robust to noise and outliers. The options provided for the algorithm are the same as those described above for K-Means.

HAC (Hierarchical Agglomerative Clustering) is a popular bottom-up clustering algorithm that merges objects to create a dendrogram, in contrast to the flat clustering algorithms presented above. The DaST platform supports the four most popular variants of HAC (single-link, complete-link, centroid-link, average-link), and the user may also tune the number of iterations and the distance method (Euclidean or Manhattan).

3.2.2 Classification. Classification identifies to which among a set of categories a new observation belongs, based on training with data that contain known data memberships.

Naïve Bayes is the simplest family of probabilistic classifiers that operate with the assumption of independence between the features. DaST supports document classification, and users are required to input only the number of documents and the document terms.

3.2.3 Information Retrieval Models. These models typically refer to a representation of documents and queries that is suitable to facilitate the retrieval of information relevant to a users need.

Boolean Model is a standard textbook representation for documents/queries that assumes a binary vector representation. Users input the query, the documents, and the scoring method (Euclidean or Manhattan).

VSM (Vector Space Model) is the most popular representation for documents/queries that assumes a real-valued vector representation. Users input the query and documents, the model variant (12 variants are supported), and the distance method (Euclidean or Manhattan).

3.2.4 Link Analysis. Link analysis refers to graph-analysis techniques used to evaluate relationships (connections/edges) between nodes (vertices) in a graph.

PageRank was introduced in the Google search engine to measure the importance of web pages. The user inputs the graph (by means of edges between vertices), the number of iterations, and the variant (with/without dampening and Google matrix usage).

HITS (Hyperlink-Induced Topic Search) is a well-known textbook algorithm that rates web pages based on the concepts of hubs and authorities; user input is the same as that of Pagerank.

3.2.5 Decentralised object location. Inspired from work in P2P networks, decentralised object location and routing currently form the backbone for most big data filesystems including GFS, HDFS, Cassandra and others.

Chord is the most famous object location protocol that is currently taught as the prominent representative of distributed hash tables (DHTs). Its key-value pair philosophy blends well with many big data system concepts. The user input involves the size of the Chord ring, the active Chord nodes, and distributed search-specific data.

3.2.6 Association rule mining. Association rule mining is a rule-based machine learning method for discovering interesting relations between transactions in large databases; it discovers interesting rules in databases using different measures of interestingness.

Apriori uses a breadth-first search strategy to count the support of itemsets and constitutes the most popular textbook algorithm for association rule mining. To create an exercise the user inputs the transaction database (up to 20 transactions – remember that the aim is exercise solving, not real-world operation), and algorithm parameters (i.e., support and confidence).

The DaST platform is regularly enriched with new algorithms and methods from the data science domain; additions are simple, due to the architectural design of the platform that allows developers to deploy new algorithms independently. In the future we plan to open-source the platform code and create a community of developers that will help the expansion of the platform and the verification of the submitted code.

3.3 Solution visualisation & step-by-step guide

As mentioned above, when selecting an algorithm/technique, the user is presented with a wizard-styled web interface to input the appropriate data and parameters for exercise creation. After the exercise creation, the user is presented with the solution of the exercise in a step-by-step manner, emphasizing the details of the algorithm. An example of such a solution is shown (in part) in Figure 2, where an exercise on the Apriori algorithm for itemset mining is solved. Please notice the stepwise execution, the color-coding and graphical elements marking the pruned and non-pruned itemsets, and the concise writeup that is meant to improve the readability and understanding of the solution.

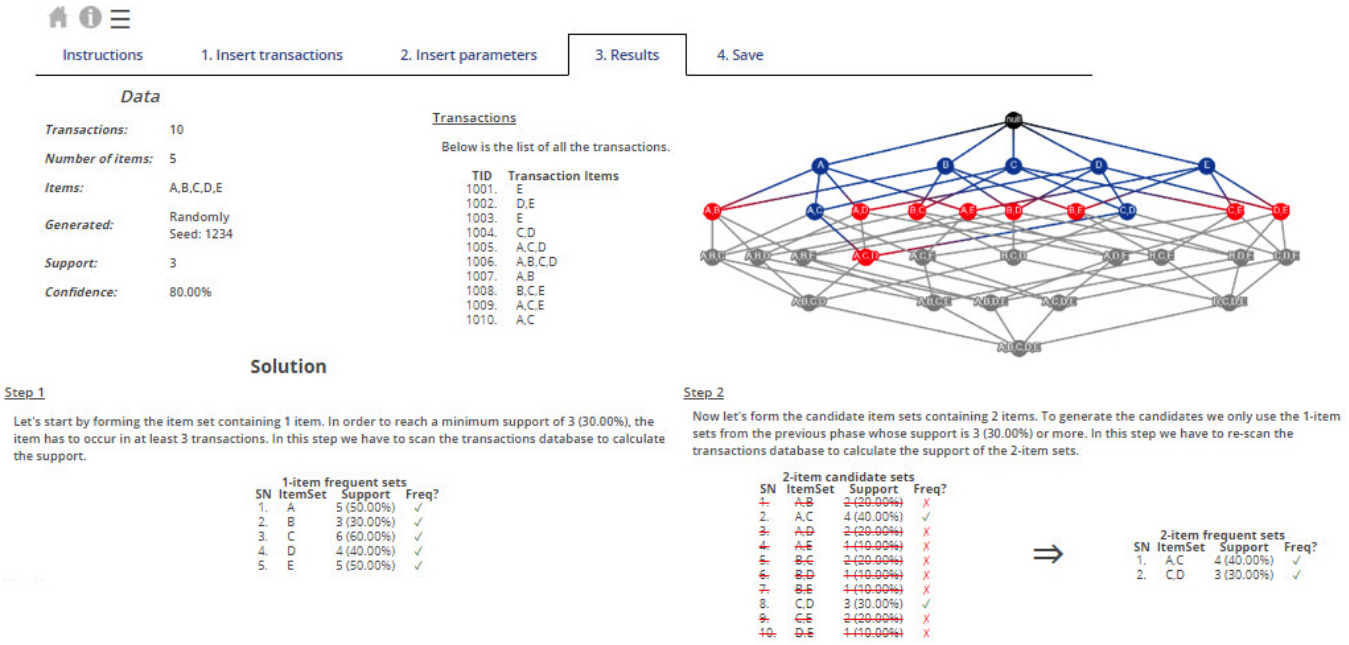


Figure 2: Solution of an itemset mining exercise (algorithm Apriori), and a graphical representation of the (in)frequent itemsets

If applicable, an appropriate visualisation of various exercise elements is included in the provided solution; for example in link analysis the user is provided with a visualisation of the input graph, while in itemset mining a color-coded latice representation of the frequent itemsets is given as a visual aid of pruned and candidate itemsets (see top right corner of Figure 2 for an example). These graphical representations of the input data and the solution space help users to get a better insight of the examined algorithm/technique and tutors to provide better explanations to students.

4 PRELIMINARY EVALUATION

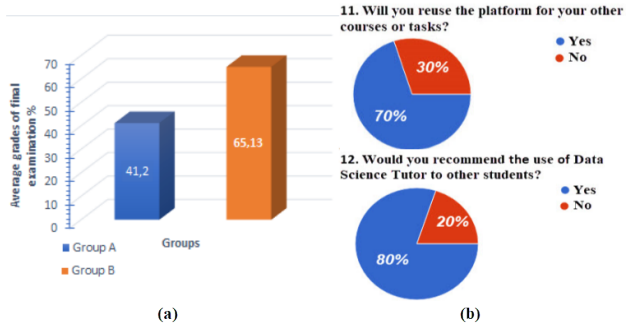
The DaST platform is online and fully functional for about one year, and we had the opportunity to make a first assessment of its acceptance by the community of tertiary education. According to user feedback we received, the platform has so far been used in a variety of courses related to the data science domain worldwide and is utilised in the context of both undergraduate and postgraduate curricula. Our assessment was carried out over three different axes: (a) qualitative and quantitative assessment from standarised anonymous questionnaires handed out from the quality assurance unit of our university, (b) student performance measurement by using a control and a test group, (c) a questionnaire survey of our own focusing on the platform specifics. The results of our preliminary assessment are reported below; we are currently in the process of performing a large scale assessment in cooperation with departments that have used our platform.

4.1 Evaluation from the quality assurance unit

Given the short life span of the platform, the time to conduct a full-scale evaluation of the learning outcomes that derive from it was limited. However, the acceptance of the DaST platform by the students of our department was encouraging. More specifically, our students evaluated the DaST platform by responding to the standarised questionnaires that are distributed by the quality assurance unit of our university and target the evaluation of the department's courses. The students' comments about the DaST platform were very positive and mostly addressed the usefulness of the platform with respect to (i) helping them "prepare for the final exams" in data science related lessons, (ii) providing them with "insights on the operation" of the respective algorithms, (iii) assisting them get non-trivial "peculiarities of the algorithms that are typically not even addressed in class", and (iv) "assessing the results and marking of tests and homeworks". One of the questions included in the questionnaires disseminated by the quality assurance unit referred specifically to the online educational material and its contribution to the better understanding of the Information Retrieval course; the students replied that the educational material was helpful and provided an overall score of 4,33 out of 5.

4.2 Students' performance report

The second axis of our assessment involved at looking into students' performance metrics, such as the achieved student grades at the final written examinations. We separated the students in one of the related courses into two different groups of ten students each; Group A had the opportunity to use the DaST platform during the semester, while Group B was not aware of the platform and did not have access to it. The chart shown in Figure 3(a) reveals the impact of the platform on student's grades as marked by the corresponding instructor; the students having used the DaST platform achieved on average 36% better score (mark range is between 0-100) in the final written examinations when compared to the students that did not use the platform. Although our sample is not big enough for extracting accurate statistical results, the particular assessment indicates that the DaST platform was beneficial to the student's performance and the learning process itself.



(a) (b)

Figure 3: (a) Average score in final exams, (b) aggregate answers to (some of) the questions

4.3 Questionnaire survey

Finally, the third assessment axis involved the design of a specialised questionnaire survey to gather information specifically concerning the DaST platform and its different dimensions (e.g., usefulness, usability, user experience). Our survey (shown in Figure 4) was performed by directly forwarding the questionnaire to users that provided email feedback to us and asking them to fill it in anonymously. The positive feedback we received, shown in part in Figure 3(b), is representative of the overall positive stance of the users' response to our questionnaire. Therefore, although an extensive user study of the DaST platform has not been performed so far, user acceptance and preliminary user feedback are encouraging.

User Profile

1. Have you ever been involved in Data Science courses such as: Machine Learning, Information Retrieval, Data Mining or Big Data?
2. Have you ever used an online platform in order to solve an exercise? If so, what was the platform? Was it free?
3. How often do you attend the the course of information retrieval?

DST evaluation

1. How useful it was the different ways of entering data?
2. How well understood is the solution of an exercise?
3. How much did the automatic solution of exercises help you understand the function of algorithms?

Suggestions for DST improvement

1. From your entire experience with the platform, what would you be more positive about?
2.From your entire experience with the platform, what would you be more negative about?
3. Give some suggestions for improving the online platform.

Figure 4: Part of the questionnaire

5 CONCLUSIONS AND FUTURE RESEARCH

The proposed DaST platform is a novel online tool that facilitates the learning and teaching of data science related algorithms/methods through *automated, step-by-step problem solving* of *user-generated* or *automatically created* exercises. It is continuously expanded with new algorithms and functionality, and has so far been warmly accepted by the data science community. Future research directions and extensions include (i) extended user studies to extract macroscopic findings concerning the learning outcomes and related targets, (ii) introducing different levels of difficulty in exercise generation, (iii) creating an online community around DaST to help development and result verification, and (iv) integration with popular LMSs such as Open eClass and Moodle.

6 ACKNOWLEDGEMENTS

This research was partially funded by project ENIRISST under grant agreement No. MIS 5027930 (co-financed by Greece and the EU through the European Regional Develop- ment Fund).

REFERENCES

[1] C.C. Aggarwal. 2015. *Data Mining: The Textbook.* Springer.
[2] D. Arnau, M. Arevalillo-Herráez, L. Puig, and J.A. González-Calero. 2013. Fundamentals of the design and the operation of an intelligent tutoring system for the learning of the arithmetical and algebraic way of solving word problems. *Computers & Education* 63 (2013).
[3] J. Berman. 2013. *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information.* Morgan Kaufmann.
[4] R.J. Brunner and E.J. Kim. 2016. Teaching Data Science. *Procedia Computer Science* 80, C (2016).
[5] W.S. Cleveland. 2001. Data science: an action plan for expanding the technical areas of the field of statistics. In *International Statistical Review.*
[6] V. Dhar. 2013. Data Science and Prediction. *Commun. ACM* 56, 12 (2013).
[7] E.N. Efthimiadis, J.M. Fernandez-Luna, J.F. Huete, and A. MacFarlane. 2011. *Teaching and Learning in Information Retrieval.* Vol. 31. Springer.
[8] G.Press. 2013. Data Science: What's The Half-Life Of A Buzzword? Forbes. https://www.forbes.com/sites/gilpress/2013/08/19/data-science-whats-the-half-life-of-a-buzzword/#47952ffa7bfd
[9] F. Grivokostopoulou, I. Perikos, and I. Hatzilygeroudis. 2016. An Educational System for Learning Search Algorithms and Automatically Assessing Student Performance. *International Journal of Artificial Intelligence in Education* 27 (2016).
[10] K. Halttunen and E. Sormunen. 2000. Learning Information Retrieval through an Educational Game. Is Gaming sufficient for learning? *Education for Information* 18, 4 (2000).
[11] S.C. Hicks and R.A. Irizarry. 2018. A Guide to Teaching Data Science. *The American Statistician* 72 (2018).
[12] K. Viswanathan Iyer. 2017. A dynamic intranet-based online-portal support for Computer Science teaching. *Education and Information Technologies* 22, 3 (2017).
[13] C.E. Izquierdo, I. López-Plata, and J.M. Moreno-Vega. 2015. Problem MetaHeuristic Solver: An educational tool aimed at studying heuristic optimization methods. *Computer Applications in Engineering Education* 23, 6 (2015).
[14] H. Keuning, B. Heeren, and J. Jeuring. 2014. Strategy-based Feedback in a Programming Tutor. In *Proceedings of the International CSERC.*
[15] J. Leskovec, A. Rajaraman, and J.D. Ullman. 2011. *Mining of Massive Datasets.* Cambridge University Press.
[16] W. Li, H. Zhou, C. Wang, H. Zhang, X. Hong, Y. Zhou, and Q. Zhang. 2019. Teaching AI Algorithms with Games Including Mahjong and FightTheLandlord on the Botzone Online Platform. In *Proceedings of the ACM Conference on CompEd.*
[17] J. Lin and C. Dyer. 2010. *Data-Intensive Text Processing with MapReduce.* Morgan & Claypool.
[18] M. Lvov, I. Chernenko, L. Shishko, and E. Kozlovsky. 2018. Mathematical Models of Supporting the Solution of the Algebra Tasks in Systems of Computer Mathematics for Educational Purposes. In *Proceedings of the 14th ICTERI. Integration, Harmonization and Knowledge Transfer Workshops.*
[19] C.D. Manning, P. Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval.* Cambridge University Press, New York, NY, USA.
[20] L. Nakayama, V.N. de Almeida, and R. Vicari. 2004. A Personalized Information Retrieval Service for an Educational Environment. In *Proceedings of the International Conference on ITS.* Springer.
[21] H. Passier, S. Stuurman, and H. Pootjes. 2014. Beautiful JavaScript: How to Guide Students to Create Good and Elegant Code. In *Proceedings of the International CSERC.*
[22] V. Putnam and C. Conati. 2019. Exploring the Need for Explainable Artificial Intelligence (XAI) in Intelligent Tutoring Systems (ITS). In *Joint Proceedings of the 24th ACM IUI Workshops.*
[23] S. Shirai, T. Fukui, K. Yoshitomi, M. Kawazoe, T. Nakahara, Y. Nakamura, K. Kato, and T. Taniguchi. 2018. Intelligent Editor for Authoring Educational Materials in Mathematics e-Learning Systems. In *Proceedings of the 6th ICMS.*
[24] P. Taele, L. Barreto, and T.A. Hammond. 2015. Maestoso: An Intelligent Educational Sketching Tool for Learning Music Theory. In *Proceedings of the 29th AAAI Conference on AI.*
[25] S. Tansley and K.M. Tolle. 2009. *The Fourth Paradigm: Data-intensive Scientific Discovery.* Microsoft Research.
[26] P. Warden. 2011. Why the term "data science" is flawed but useful. O'Reilly Radar. http://radar.oreilly.com/2011/05/data-science-terminology.html
[27] S. Willms. 2003. Visualizing a User Model for Educational Adaptive Information Retrieval. In *Proceedings of the International Conference on UM.* Springer.