

# Supporting Decisions Using Educational Data Analysis

Laura O. Moraes

[laura.moraes@coppe.ufrj.br](mailto:laura.moraes@coppe.ufrj.br)

Programa de Engenharia de Sistemas e Computação,  
COPPE/UFRJ and Colégio Pedro II

Carla Delgado

Instituto de Computação, UFRJ

[carla@ic.ufrj.br](mailto:carla@ic.ufrj.br)

Carlos Eduardo Pedreira

[pedreira56@gmail.com](mailto:pedreira56@gmail.com)

Programa de Engenharia de Sistemas e Computação,  
COPPE/UFRJ

João Pedro Freire

[joaofreire@poli.ufrj.br](mailto:joaofreire@poli.ufrj.br)

Bacharelado em Ciências Matemáticas e da Terra, UFRJ

## ABSTRACT

We present the Machine Teaching, an online learning environment with two main goals: (1) supporting student practicing and exercise marking; and most important, (2) collecting data on students' knowledge while they progress. Machine teaching was key to bringing programming courses to online learning during the 2020 pandemic, helping educators provide a safe and smooth online practice environment for students and helping them to master programming skills in early stages of their bachelor's degree studies, a skill that increases the possibilities for immediate job placement. In addition, the educational data collected are mined and used to support short- and long-term pedagogical decision-making, allowing for a quick feedback and enabling material adaptations for the classes offered in the remote mode.

## KEYWORDS

educational data mining, computer science education, learning analytics

## 1 INTRODUCTION

In regular face-to-face classrooms, educators perform "knowledge discovery" (in the educational data mining sense) everyday, by observing student behavior and outcomes in assignments and tests. From their observations, they can determine which concepts students have mastered or are still struggling with and consequently adapt teaching strategies. In distance learning this observation is limited and, therefore, we need enhanced tools to support this task. Also, student log analysis can provide insights to improve teaching and learning strategies, course curricula and other sense-making processes. However, for it to be useful, this amount of data have to be presented in a way for stakeholders to act on it. Recent reports in learning analytics pointed out that we need to approximate the already rich published research to actual deployment and practice [4, 10, 11].

We present an online learning environment whose main objective is to collect data on students' knowledge during the programming learning process to support short-term and long-term pedagogical decision-making. In this way, weak points and opportunities for improvement in the teaching-learning process can

be identified and worked on. Just as important as the main objective, which is strategic, the environment has also an operational objective: supporting student practicing and exercise marking.

*Machine Teaching* is a web system to support introductory Computing learning that has been developed at the Federal University of Rio de Janeiro and used as a support tool for introductory computing courses at this University since 2018. Secondary objectives of this environment are: to provide immediate feedback during activities, which we consider primarily a stimulus factor for students in their practices, and to partially automate the analysis of student codes, facilitating the work of monitoring the generation of feedback from teachers and tutors. It is also worth mentioning that the adoption of interactive content for learning in classrooms and in self-learning environments has been shown to engage students [1, 6, 7], improve learning, decrease dropout and failure rate [1, 3, 5] and increase self-confidence, especially in female students [5].

Besides supporting learning and teaching of programming in UFRJ and very soon to further institutions, the Machine Teaching project also has a very ambitious research agenda: investigate if the data collected from learning environments used in the introduction to programming courses can provide decision support for different stakeholders. Our target users are educators (lecturers and teaching assistants) and students and our research questions are:

- (1) RQ1: Can we improve educators awareness on students difficulties?
- (2) RQ2: Can we drive students to adapt their studying strategies by showing to them their weak points?

The main impacts already achieved by this project are: (1) smooth and quick transition of introductory programming courses to online set up, making possible for students to conduct their practices alone and in their own settings; (2) helping students to acquire programming skills which increased their possibilities for immediate job placement (a demand of many students as a consequence of the financial crises faced in the past year); (3) providing EDM results and strategies to use this data to make teaching-learning process more efficient and effective, and meaningful.

## 2 MACHINE TEACHING: DATA ACQUISITION SYSTEM AND USER INTERACTION

This section presents the teaching methodology used to design and implement the Machine Teaching learning environment. Then, we present the web system developed to acquire student data and interact with students and educators, presenting its main features.

In: XX Workshop de Ferramentas e Aplicações (WFA 2021), Minas Gerais, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2021.

© 2021 SBC – Sociedade Brasileira de Computação.

ISSN 2596-1683

## 2.1 Teaching Methodology

The introductory programming course currently offered by the Computer Science Department of the Universidade Federal do Rio de Janeiro (DCC/UFRJ) for students of non-computing careers aims to develop the skills for building readable and modular Python programs. This methodology is based on the structured imperative approach, emphasizing problem-solving, procedural decomposition and basic skill mastery. This didactic proposal inverts the usual order of structured imperative teaching that usually starts with a complete program structure and user interaction such as “*print*”, “*input*”, and “*\_\_main\_\_*” statements. The proposal differential is that most of the emphasis consists of building concise code modules, leaving the user interaction mechanisms to the end of the course (when the student already mastered the basics). The adopted approach provides the student with an orientation towards the most abstract cognitive tasks of program design and construction since the beginning of the learning process. The detailed methodology description can be seen in [2].

## 2.2 User Interaction

This section presents the main Machine Teaching dashboards and how the user interact with the system. In the integrated development environment, the students are presented with a problem, and they should write the expected answer in a free-text coding format. For each exercise, a test case function generator was defined to correct the results. The students get feedback every time they submit an answer, and they can see whether they passed or failed a unit test case. If they get all of them correct, the task is considered done, and the student may move on to another problem. The system saves a state every time a student submits an answer.

Students have access to a dashboard where they can visualize their completion statistics and compare themselves to their peers, as shown in Fig. 1. This dashboard shows how long the student took to finish the class, the number of errors, and each class’ and total progress.

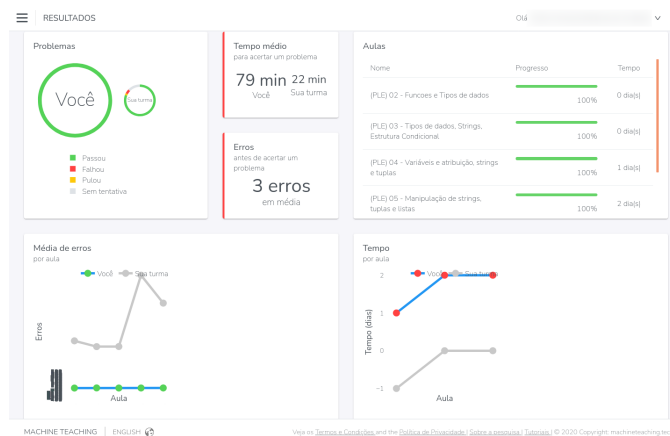


Figure 1: Student dashboard

The Machine Teaching system provides three interfaces for the professors to understand students’ submissions at class or individual levels. Fig. 2 shows one of them and presents a class overview

for educators. In this dashboard, educators can understand which students do not finish the exercises on time and how many tries they attempt before getting the exercise correctly.

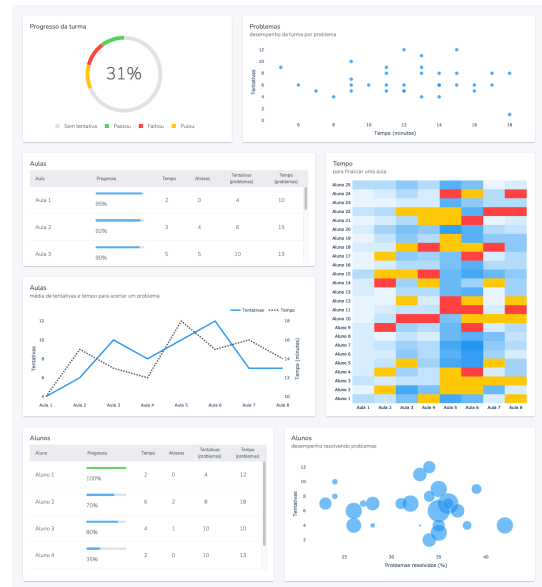


Figure 2: Class dashboard. Time spent and quantity of attempts per problem and student are displayed in this interface.

## 3 RESULTS

The Machine Teaching web system was developed to support data-driven decision making by educators and students. The system architecture and data collection methodology [8] were already shared with the academic community in international conferences. Also, the data collected using the system has already shown to be useful for educational data mining research [9, 12], being used to cluster students responses in topics and in performance prediction models by one of our partners.

In this section, we validate if the system’s dashboards can be useful for students and educators to support their decision making. We surveyed students and professors about the system usability and their decision-making process when using the system. Both surveys were done in remote classes during 2020, after the COVID-19 pandemic had started. Participation was anonymous and voluntary, and the survey was conducted on Google Forms using a Likert scale with five values ranging from “Disagree completely” to “Agree completely”.

### 3.1 System Usability Survey

We investigated the user interface satisfaction: subjects found the Machine Teaching system easy to use; they could find the exercises that had to be done, and the exercises that they had already done. We asked the following questions: (1) The system has a friendly interface; (2) It was easy for me to find past exercises to study; (3) It was easy for me to find this week’s exercises.

In total, 267 students replied. Fig. 3 shows the students' responses concerning interface satisfaction. In general, they were able to navigate through the system and find past and current exercises. Current exercises had a higher median than past exercises since the system was projected to present the most recent exercises on top. Friendly interface perception had a higher dispersion than the other two questions, but it still received a median of 4 out of 5 on the Likert scale.

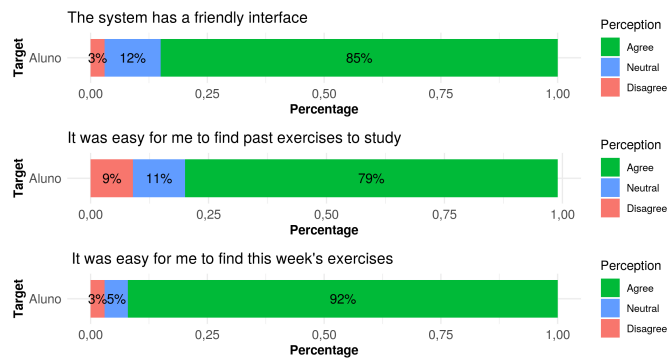


Figure 3: Interface satisfaction

### 3.2 Decision-making Support Survey

To answer our research questions and start investigating if it is possible to impact students' and educators' decision making, we asked them if the information displayed in the dashboards helped them in organizing the activities and obtaining awareness of time spent in each activity. We asked the following questions:

- (1) About time organization:
  - (a) **For students:** the dashboard helped me organize my study time
  - (b) **For educators:** the system helped me to have a better view of the effort and/or time that students spend doing the exercises
- (2) About content organization:
  - (a) **For students:** the dashboard helped me to guide the content I should study by pinpointing the most common mistakes
  - (b) **For educators:** the system helped me to identify content that is more difficult for students

Figs. 4 and 5 show their perception about these subjects. Most of the students is not using the information provided by the dashboard to adjust the time dedicated to studying the subject and about 40% of them use the most common mistakes to guide their study. We plan to release a new and simplified dashboard version for students. The current version combines a lot of information into a single screen. We are designing the new one to provide more direct tips to students on how to improve their performance.

Educators received the most gain by using the system. Before it was deployed, the students used to send files by email containing the code provided as answer. Educators had no way of knowing how much time each student spent in each exercise and which ones they had more difficulty in getting the correct answer. This can

be verified by the educators evaluation of both time and content organization questions. We released the educators dashboard (Fig. 2) at the beginning of the 2nd semester of 2020, improving the positive perception from 17% to 75%.

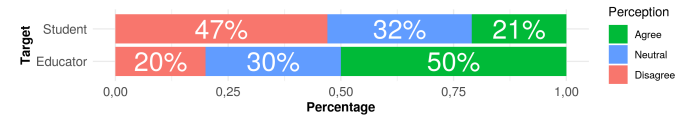


Figure 4: Student and educators perception on time organization using information provided by the dashboard

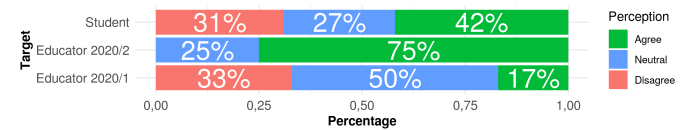


Figure 5: Student and educators perception on content organization using information provided by the dashboard

## 4 CONCLUSION

This paper presented a learning environment currently in use at the Universidade Federal do Rio de Janeiro in Introduction to Computing classes. During the year 2020, UFRJ has become completely remote due to the COVID-19 pandemic. Having an already tested and deployed system tailored for the teaching methodology used in the introduction to programming classes helped in the transition from face-to-face classes to remote classes.

We designed and implemented a learning environment based on a teaching methodology, inserting educators and students into the designing loop. We surveyed educators and students from the previous two semesters about their perceptions of the system. The educators survey showed that they perceive the system as a useful tool to gain awareness about their student performance and difficulties (RQ1). The student survey showed that, even though students perceived the system as helpful for learning, they did not use it to optimize their studying strategies (RQ2). This is a point that should be further investigated, so that we can better understand how educational data mining could positively influence students self assessment of their knowledge and adapt their studying routines. In spite of that, students' experience using the system was positive overall, and we could identify improvement points, especially on the time management topic.

Dashboards for educators were particularly helpful for marking students exercises. This was very important in the remote configuration of the courses, as evaluation instruments and schemes had to be dramatically different from what it used to be in classroom. Recently, dashboards gathering mined data from the crossing of students, educators, classes, exercises, and content were made available to educators and course administrators, but we were not yet able to investigate their impact in decision making. Our preliminary analysis of these data provides us with evidence that it can significantly impact planning, provision and execution of the courses

each semester, as well as content and exercise organization and didactic strategies adapted to the remote scenario. As future work, we plan to investigate and document strategies for the usage of EDM in the support and improvement of educational settings.

#### 4.1 Social Impact

Due to the COVID-19 pandemic, an abrupt transition from presence to online learning settings had to be faced by educators and students. Machine Teaching was of great help to support this transition as it provided the students the possibility of conducting their practices alone and in their own settings. We consider this a considerable social impact, not only because of the importance of UFRJ to education in Brazil, but also because CS1 courses are strategic for students who needed to prematurely start their professional careers because of family financial losses suffered as side effects of the crises around COVID-19 pandemic. At the present moment, the built web-system is being used by 15 classes and 500 students. In total, more than 1600 students have used the system in the past three years. We are working to make Machine Teaching available to be used in other public education institutions. This requires, besides operational upgrades, a close contact with educators who are willing to use it. Our expectation is to have 1 further institution besides UFRJ using Machine Teaching still this year, and at least 2 next year.

Furthermore, strategies for using educational data mining (EDM) to support and improve educational environments have the potential to make educational processes more efficient and effective, a demand perceived not only by UFRJ and Brazil, but throughout the world.

#### 4.2 Schedule

For the next steps, we plan on expanding the Machine Teaching to be used by other institutions besides UFRJ. The system already has an english version and will be used by Universidad de Alcalá de Henares from Spain in September. The system is already prepared and no adaptations will have to be done in this setting.

We also submitted a joint proposal for FAPERJ Woman in Science project together with the Minerva UFRJ group, Colegio Pedro II and Secretaria Municipal de Educação de Tanguá to use the system to promote computational thinking for woman in schools. In this case, the system will have to be adapted to support other programming languages besides Python. We foresee to do this step still this year.

These new experiences will also provide us with data to improve the dashboards and support decision-making processes. Therefore, we plan on releasing at least two new dashboard versions this year and one paper with exploring our findings.

#### REFERENCES

- [1] Benotti, L. et al. 2018. The Effect of a Web-based Coding Tool with Automatic Feedback on Students' Performance and Perceptions. In *Proc. 49th ACM Tech. Symp. Comp. Sci. Educ. (SIGCSE '18)*. Baltimore, Maryland, USA, 2–7. <https://doi.org/10.1145/3159450.3159579> <http://doi.acm.org/10.1145/3159450.3159579>.
- [2] Delgado, C. et al. 2016. The teaching of functions as the first step to learn imperative programming. In *Anais do Workshop sobre Educação em Computação (WEI)*. Sociedade Brasileira de Computação - SBC, 388–397. <https://doi.org/10.5753/wei.2016.9683> doi: 10.5753/wei.2016.9683.
- [3] Ihanola, P. et al. 2015. Educational Data Mining and Learning Analytics in Programming: Literature Review and Case Studies. In *Proc. 2015 ITiCSE Work.*

*Group Rep. (ITiCSE-WGR '15)*. Vilnius, Lithuania, 41–63. <https://doi.org/10.1145/2858796.2858798> doi: 10.1145/2858796.2858798.

- [4] Jelena Jovanović, Shane Dawson, Srećko Joksimović, and George Siemens. 2020. Supporting actionable intelligence: reframing the analysis of observed study strategies. In *Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 161–170. <https://doi.org/10.1145/3375462.3375474>
- [5] Amruth N. Kumar. 2008. The Effect of Using Problem-Solving Software Tutors on the Self-Confidence of Female Students. In *Proc. 39th SIGCSE Tech. Symp. Comput. Sci. Educ. (SIGCSE '08)*. Portland, OR, USA, 523–527. <https://doi.org/10.1145/1352135.1352309>
- [6] Latham, A. et al. 2012. A conversational intelligent tutoring system to automatically predict learning styles. *Comput. & Educ.* 59, 1 (Aug. 2012), 95 – 109. <http://www.sciencedirect.com/science/article/pii/S0360131511002648> <https://doi.org/10.1016/j.compedu.2011.11.001>.
- [7] Richard Lobb and Jenny Harlow. 2016. Coderunner: A Tool for Assessing Computer Programming Skills. *ACM Inroads* 7, 1 (Feb. 2016), 47–51. <https://doi.org/10.1145/2810041> <https://doi.org/10.1145/2810041>
- [8] Laura O. Moraes and Carlos Eduardo Pedreira. 2020. Designing an Intelligent Tutoring System Across Multiple Classes. In *4th Educational Data Mining in Computer Science Education Workshop (Virtual)*.
- [9] Laura O. Moraes and Carlos Eduardo Pedreira. 2021. Clustering Introductory Computer Science Exercises Using Topic Modeling Methods. *IEEE Trans. Learn. Technol.* (2021). <https://doi.org/10.1109/TLT.2021.3056907>
- [10] Abelardo Pardo, Kathryn Bartimote, Simon Buckingham Shum, Shane Dawson, Jing Gao, Dragan Gašević, Steve Leichtweis, Danny Liu, Roberto Martínez-Maldonado, Negin Mirriahi, Adon Christian Michael Moskal, Jurgen Schulte, George Siemens, and Lorenzo Vigentini. 2018. OnTask: Delivering Data-Informed, Personalized Learning Support Actions. *Learning Analytics* 5, 3 (Dec. 2018), 235–249. <https://doi.org/10.18608/jla.2018.53.15>
- [11] Carolyn P. Rosé, Elizabeth A. McLaughlin, Ran Liu, and Kenneth R. Koedinger. 2019. Explanatory learner models: Why machine learning (alone) is not the answer. *British Journal of Educational Technology* 50, 6 (Nov. 2019), 2943–2958. <https://doi.org/10.1111/bjet.12858>
- [12] Chunpai Wang, Shaghayegh Sahebi, Siqian Zhao, Peter Brusilovsky, and Laura O. Moraes. 2021. Knowledge Tracing for Complex Problem Solving: Granular Rank-Based Tensor Factorization. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21)*. Association for Computing Machinery, New York, NY, USA, 179–188. <https://doi.org/10.1145/3450613.3456831>

#### A TEAM AND RESOURCES

Our team currently has 4 members: **Laura O. Moraes** started this project during her PhD at UFRJ (supported by CNPq under Grant 141089/2016-4). She designed and implemented the platform and conducted the first data analysis (CV Lattes: 3138892444406479). **Carla Delgado** is co-responsible for the construction and management of the teaching methodology and material that is used in the Machine Teaching platform and in the programming courses offered by the Computing Institute to students of practically all undergraduate courses at UFRJ in STEM fields (CV Lattes: 3831909651244142). **João Pedro Freire** is a CNPq Scientific Initiation student in this project and is a Bachelor student of Science in Mathematics and Earth. As a student, he was a user of the Machine Teaching system and volunteered to work in the system development (CV Lattes: 6330024986448441). **Carlos Eduardo Pedreira** was Laura's advisor during her PhD. Presently, he is a professor and the head of the AI sector at PESC/COPPE (CV Lattes: 2718664296804955, supported in part by FAPERJ under Grant E26-200.840/2021-CNE, by CAPES under Grant PROEX - 1201036, and CNPq under Grant 306258/2019-6).

#### B ONLINE RESOURCES

- (1) Machine Teaching web-page: [www.machineteaching.tech](http://www.machineteaching.tech)
- (2) Our presentation at Festival do Conhecimento da UFRJ 2021: [https://www.youtube.com/watch?v=7GCnZ\\_WyH6U](https://www.youtube.com/watch?v=7GCnZ_WyH6U)
- (3) Python UFRJ web-page: <https://dcc.ufrj.br/~pythonufrj/>