Gamifying the Classroom for the Acquisition of Skills Associated with Machine Learning: A Two-Year Case Study

Chapter · November 2022

DOI: 10.1007/978-3-031-18409-3_22

6 authors, including:

Antonio Manuel Durán-Rosal
University of Cordoba (Spain)

David Guijo-Rubio
University of Cordoba (Spain)

Víctor Manuel Vargas Yun
University of Cordoba (Spain)

Pedro Antonio Gutiérrez
University of Cordoba (Spain)

Gamifying the classroom for the acquisition of skills associated with Machine Learning: a two-year case study

Antonio M. Durán-Rosal, David Guijo-Rubio, Víctor M. Vargas, Antonio M. Gómez-Orellana, Pedro A. Gutiérrez, Juan C. Fernández

Abstract Machine learning (ML) is the field of science that combines knowledge from artificial intelligence, statistics and mathematics intending to give computers the ability to learn from data without being explicitly programmed to do so. It falls under the umbrella of Data Science and is usually developed by Computer Engineers becoming what is known as Data Scientists. Developing the necessary competences in this field is not a trivial task, and applying innovative methodologies such as gamification can smooth the initial learning curve. In this context, communities offering platforms for open competitions such as *Kaggle* can be used as a motivating element. The main objective of this work is to gamify the classroom with the idea of providing students with valuable hands-on experience by means of addressing a real problem, as well as the possibility to cooperate and compete simultaneously to acquire ML competences. The innovative teaching experience carried out during two years meant a great motivation, an improvement of the learning capacity and a continuous recycling of knowledge to which Computer Engineers are faced to.

Antonio M. Durán-Rosal
Dept. of Quantitative Methods, Universidad Loyola Andalucía, e-mail: amduran@uloyola.es

David Guijo-Rubio
Dept. of Computer Science and Numerical Analysis, University of Córdoba, e-mail: dguijo@uco.es

Víctor M. Vargas
Dept. of Computer Science and Numerical Analysis, University of Córdoba, e-mail: vvargas@uco.es

Antonio M. Gómez-Orellana
Dept. of Computer Science and Numerical Analysis, University of Córdoba, e-mail: am.gomez@uco.es

Pedro A. Gutiérrez
Dept. of Computer Science and Numerical Analysis, University of Córdoba, e-mail: pagutierrez@uco.es

Juan C. Fernández
Dept. of Computer Science and Numerical Analysis, University of Córdoba, e-mail: jfcaballero@uco.es

1 Introduction

Machine Learning (ML) [2] allows computers to learn from data without having to be programmed to do so. This area, which combines knowledge of statistics, mathematics and artificial intelligence, is part of what is known as Data Science (DS) [1]. DS is based on developing scientific methods, processes and systems to extract knowledge from existing datasets to identify, analyse, understand or even improve current processes in these fields. DS and ML are gaining special interest in recent years due to their great interdisciplinarity. Numerous fields are benefiting from the advances in these areas, such as meteorology [13] or renewable energies [11], among others. Furthermore, it is common to find organisational processes in any sector and industry applying these techniques [4]. Moreover, governments have moved toward creating governmental strategies around Artificial Intelligence (AI). It is not difficult to find promising numbers on DS and ML collected by different organisations, e.g. *LinkedIn* [9], *Glassdoor* [16], the *US Bureau of Labor Statistics* [5] or *IBM* [20].

A person who works with DS and ML models is known as a Data Scientist. Their main objective is to understand and analyse the fundamental phenomena that happen, using techniques and theories drawn from various fields. Specifically, the Data Scientist profile is related to knowledge in mathematics, statistics, programming languages, and practical experience in analysing real data and elaborating predictive models. Until now, this discipline has been developed through different professional profiles, such as mathematicians or statisticians, although in recent years, the profile of Computer Engineer has been the most appropriate, given the knowledge and skills in the aforementioned areas.

The critical interdisciplinarity and the need for automatic and efficient information processing make the profession of Data Scientist one of the most sought-after jobs today. However, as it is a theoretical and practical area at the same time, it is a handicap for students to acquire the necessary competences at the university stage. Hence, universities must provide not only theoretical but also practical knowledge, which is difficult due to the breadth of knowledge to be satisfied in the BSc in Computer Engineering. Recently, one of the ways to make lessons more practical and entertaining is gamification. It is defined as the use of game design elements in non-game contexts [10]. Furthermore, it has received increased attention and interest in academia and practice due to its motivational power [26]. Competitions between students, including points, leader-boards, rewards and prizes, are considered one of the ways to include gamification in the classroom. In this context, platforms such as *Kaggle*[1], founded in 2010, have emerged with the primary objective of connecting companies whose main function is to provide data to solve a given problem through competitions, with researchers and Data Scientists, who bring solutions to these problems through existing techniques in DS and, more specifically, ML.

Motivated by previous works of teaching DS and ML to students of the BSc Computer Engineering [3, 25, 15, 22], the major objective pursued with this work was

[1] https://www.kaggle.com

to complement the knowledge, mainly theoretical, taught in the different subjects of this Degree, with a completely practical experience, repeating the process during two years in order to extract more consistent conclusions. With such practical experience, the students of the different subjects benefited from professionals trained in this type of task, given that this is the field of research of the professors associated with these subjects [17]. In this way, the students could apply a wide range of techniques presented during the development of the subjects and, thus, face the complexity of a real problem. Besides, and transversely to the project, another competence developed was the ability to work remotely.

For the reasons above, *Kaggle* was used as an Information and Communication Technologies (ICT) tool in two subjects of the BSc in Computer Engineering offered by the University of Córdoba (Spain). In this way, the primary theoretical learning received during the BSc is complemented and applied to an eminently practical problem through a competition via Kaggle. The underlying idea was that competition, considered an element of gamification [18, 6], would increase the students' motivation to find the best solution to the proposed problem. Hence, the students would face the standard workflow that is carried out when tackling a problem of this type, consisting of preprocessing of the databases, design and training of the models, validation of the models obtained, adjustment of parameters, and finally, evaluation of the predictive models. This innovation experience was part of the development of two Teaching Innovation Projects (TIPs) developed by the authors of this work. Both TIPs were awarded by the University of Córdoba according to one of the priority lines of action established in the modality: transfer of theoretical knowledge into practice.

Before the *Kaggle* competition, the teaching staff associated with the subject tutored lent their experience in the field of ML and DS to address a real problem and provided guidance on how to tackle the competition. The main objective of these workshops was to provide specific training, on the *python*[2] programming language and the *scikit-learn* library [23]. It is important to highlight the use of these tools in ML: *python* is the most widely used language for these tasks due to its enormous versatility, learning curve, and ease of use, among other characteristics [27]. In addition, the students used only free software throughout the TIPs, joining the line of action in favour of it set by the Conference of Rectors of Spanish Universities and with the support and collaboration of the *Free Software Group*[3] of the University of Córdoba. On the other hand, *InClass Competitions* service in *Kaggle* was used. This service is offered free to the teaching and research community for its use with students. Its main objective is to make it easier for DS faculty to run a competition exclusively for students so that the level is tailored to them, and the remainder of the *Kaggle* community does not have access. The results of the competitions were very satisfactory, not only because of the increase in knowledge on the part of the students but also because of the high level of motivation with which they attended the

[2] https://www.python.org

[3] https://www.uco.es/aulasoftwarelibre/

workshops. Besides, the level and rate of participation in the competitions increased during the second year.

2 Gamification in Data Science and Machine Learning

As mentioned above, gamification includes game elements in non-game contexts. This supports motivation and performance in skills such as problem-solving, collaboration and communication, which are essential competencies in the Data Scientist profession. In the last decade, gamification has supported learning in many application areas. Of particular interest are ML and DS [19], being Computer Engineering one of the areas with the most prominent application of gamification [21].

Points, leaderships, avatars, prizes and rewards are the most common elements in gamification experiences, so competitions are the ultimate way to bring all these components together. In this regard, a game-based competition for teaching artificial intelligence was proposed in [7]. For this purpose, the Nine Men's Morris game was selected to develop an artificial player using heuristic search, optimisation and ML concepts. The competition consisted of a round-robin tournament where each player had to play against the others. The results confirmed that the experience was a success. The students reported an improvement in their skills and declared an increased interest in the course topics, and in AI in general. The effectiveness of the learning approaches adopted was also confirmed.

More important, and as aforementioned, *Kaggle* is the platform par excellence for carrying out ML problem-solving competitions. Related to this point, a methodology for teaching ML using a game-like *Kaggle* competition to make the learning more engaging and fun was proposed in [8]. Specifically, the methodology consisted of seven steps, including 1) teaching knowledge, 2) briefing competition guidelines, 3) giving sample solutions, 4) registering for the competition, 5) submitting and discussing the results, 6) reflecting on the learning experience, and 7) scoring the teams. As a result, students were motivated to look for different ways to solve problems using ML techniques, and also to improve their understanding on how to apply the theoretical concepts and algorithms to real-world problems.

In [24], the authors conducted a study to determine whether the use of *Kaggle InClass* enhances problem learning in DS and ML. Sixty-one students were divided into two groups for regression and classification problems, building prediction models individually for 16 days, and forming groups for other 7 days. The students that participated in the competition performed better in the exam. Moreover, they found the competition exciting and valuable for their learning in the course.

More recently, García-Algarra [12] proposed a methodology for teaching ML concepts in non-STEM (Science, Technology, Engineering and Maths) areas. The course went through different phases to teach a group of people without experience in this field or programming about related concepts. In the last phase of the course, there was a *Kaggle*-like competition forming groups between people who did not know each other, which was worth 10% of the course grade, with the winning group getting an additional 10%. This provoked a high degree of interest and effort, which contributed to improve the understanding of the concepts taught regarding ML.

3 Development of the innovation experience

This work was focused on the subjects "Introduction to Machine Learning" and "Introduction to Computational Modelling" of the 3rd and 4th year of the BSc Computer Engineering of the University of Córdoba, respectively, during the academic years 2017/2018 and 2018/2019.

3.1 Objectives

The main objectives associated with the implementation of the two TIPs were the following:

1. To improve the acquisition of practical aspects of the specific competences belonging to the Data Scientist professional profile.
2. To instruct students in the use of the *Kaggle* platform as a tool to learn the applicability of ML concepts in a fun and entertaining way.
3. To broaden the knowledge of programming languages and DS.
4. To combine a cooperative and competitive environment for problem-solving.
5. To foster an atmosphere of cooperation and competitiveness at the same time among the students.
6. To confront students with the complexity of solving real-world problems using ML techniques.

3.2 Materials and methods

The above objectives were carried out through the activities described below, which were common to the two TIPs.

Activity 1. Development of the competition. The first step was to find a real-world problem resulting attractive for the students. For this, a real significant wave height prediction problem in the Gulf of Alaska (United States) was selected. This problem, focused on the need to anticipate events in the natural environment and in the field of renewable energies, had been previously tackled by members of the research group, thus, its difficulty could be adapted to be affordable by the students of the subjects. The dataset was built using 18 input variables related to meteorological observations, which were both obtained from sensors installed on buoys and also from a reanalysis model. For the significant wave height prediction, the continuous variable was discretised into 4 categories according to the significant wave height: *low*, *medium*, *moderate* and *very high*. Then, the problem was adapted according to the *Kaggle* platform, which requires three datasets: 1) the training dataset for building the models, 2) the public testing dataset to validate them, and 3) the private testing dataset to check the fairness of the public validation. For these two validations, the students had to upload the predictions obtained by their developed models to the *Kaggle* platform, which automatically generates a public and a private ranking. Note that the private ranking on a previously unseen testing dataset is only displayed at the end of the competition.

Activity 2. Multi-session practice. In order to introduce the students to the *Kaggle* platform, a 3-session practice was carried out. In these sessions, it was explained how the platform works and the real problem they were going to tackle. For this, the dataset detailed in Activity 1 is considered as well as several case studies were proposed as examples, providing their source code, so that they could test and practice with different ML methods, favouring the understanding of each of them.

Furthermore, the work performed by the students in this Activity was going to be evaluated using the following criteria: 40% of the mark was scored based on the private ranking as it measures the quality of the developed model, and the remaining 60% was scored according to the workflow implemented by the students to obtain the models: preprocessing and visualisation of the data, analysis of data (outliers, extreme values and correlated variables), feature's selection, etc. After the first session, the competition was launched, closing one week after the last practical session.

Activity 3. Working team for tackling complex competitions. In order to tackle complex competitions on *Kaggle*, a working team was established between teachers and students. The main goal of this working team was to cooperate for the benefit of the team and to compete against other teams. Meetings were planned frequently during tutoring hours, both face-to-face and virtually.

Activity 4. ML training. Apart from the multi-session practice, a series of workshops on ML were given by the teachers in collaboration with the *Free Software Group* of the University of Córdoba. The main objective was to complement the training received by the multi-session practice (Activity 2). Up to now, the students were instructed to create predictive models using Weka [14]. Therefore, it was a good opportunity to teach the *python* programming language together with the *scikit-learn* library. Two sessions were held each year[4].

Activity 5. Evaluation of the TIPs. For assessing the practical knowledge acquired by the students of both subjects, two voluntary test-type questionnaires were carried out: the first one before the start of the multi-session practice, and the second one after the end of the competition. Both questionnaires included the same 20 questions: 16 for assessing the students' level regarding ML, and the remaining 4 regarding self-assessment questions, on a Likert scale, about the *Kaggle* platform, *python* and *scikit-learn*.

4 Results

In this section, the results of both TIPs are presented, compared and discussed. These results are divided between those obtained through Activity 5, reflecting the evaluation of the TIPs through the questionnaires completed by the students, and those obtained from the teaching innovation experience, through the considerations of the work team.

[4] Please, check the agenda of each tutorial in the GitHub repositories https://github.com/ayrna/tutorial-scikit-learn-asl and https://github.com/ayrna/taller-sklearn-asl-2019, for the first and the second TIP, respectively.

4.1 Assessment questionnaires

The forms were completed in two phases, before carrying out the set of activities and after their completion. Moreover, the form was divided into two parts, the first part, consisted of 16 multiple-choice questions, to obtain the students' level and 4 self-assessment Likert scale questions focused on the main tools used in the project. First, the results of the multiple-choice questions are analysed before and after the activity for both years considered. After that, the self-assessment results are presented, again for both years and before and after the sessions.

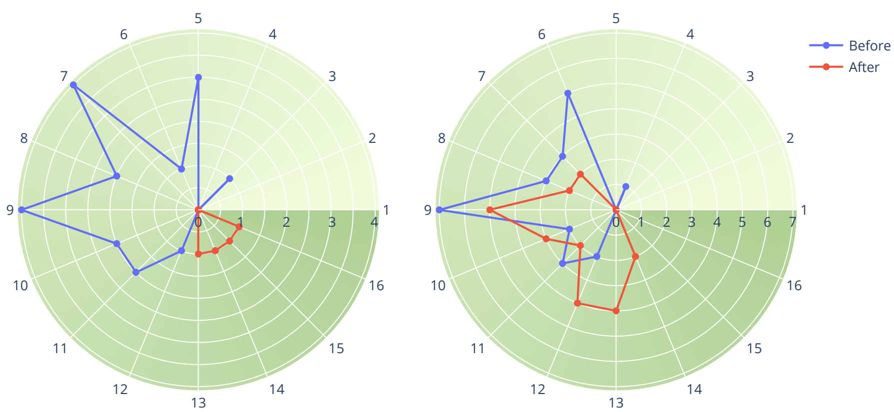

Fig. 1 Results obtained in the multiple-choice questions before and after the sessions for the first (left) and the second (right) years. Outer values represent the total score obtained (from 1 to 16) and the inner values indicate the number of students who obtained each of these scores.

In the first year, a total of 19 students took part in the questionnaires. While all of them did the initial tests, only a small part of them completed the final test. The completion of the test form before the sessions allowed us to know the initial level of the students in relation to ML and DS tasks. Each of the questions of the first part where evaluated with 1 point, thus reaching a maximum of 16 points. The average score obtained by the students before the sessions was 7.9 out of 16, which is 4.9 out of 10. Regarding the test performed at the end of the workshops, the average score raised to 14.5 out of 16, which is 9 out of 10. Also, the range of scores went from 3-12 to 13-16. These results are shown in a graphical way in the left chart of Fig. 1, where the outer values represent the total score obtained (from 1 to 16) and the inner values indicate the number of students who obtained each of these scores. For instance, two students obtained a score of 11 in the questionnaire carried out before the sessions, whereas one student obtained a score of 15 in the questionnaire performed after the sessions.

In the second year, the total number of students that took part in the activity raised to 26. The evaluation method was the same as the one described for the first year. For the first part, the average score obtained by the students before taking part in this project was 8.4 out of 16, which is 5.3 out of 10. Then, after their participation in the organised sessions, their average score raised to 10.6 out of 16, which is 6.6 out of 10. Also, the interval of scores was shifted to the upper part of the ratings, given that in the initial test it was 4-12 and, in the end, it was 7-14. These results, which show that the students improved their knowledge and ML skills, are graphically shown in the right chart of Fig. 1.

Both charts represented in Fig. 1 show that the scores of the questionnaires taken before the activity are concentrated on the low-medium part of the rating range while, those taken after the sessions, they are mostly located on the medium-high part. In fact, in the first year, there is no overlapping between both scores, given that the final results are only in the 13-16 range, i.e. the 4 students obtained excellent scores. For the case of the second year, there is an important overlapping between both sets of scores, but the scores obtained after the activity are clearly biased towards the highest scores.

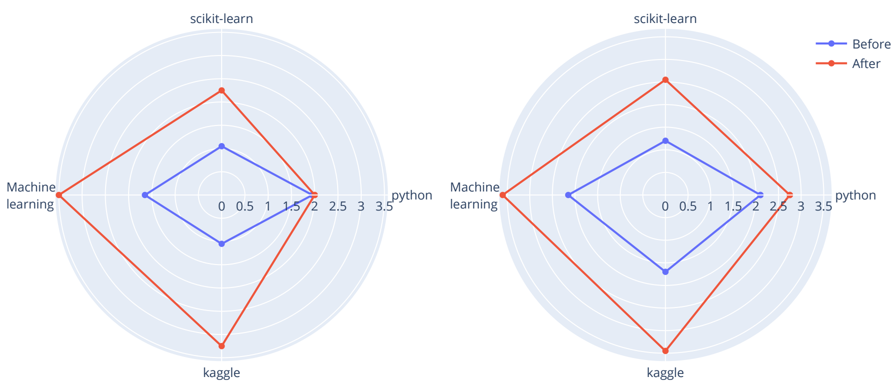

Fig. 2 Results obtained in the self-assessment questions before and after the sessions for the first (left) and the second (right) years. The magnitude of the chart represents the mean score obtained by the students on each of the skills (outer values) following the Likert scale (from 0 to 4.

Concerning the second part of the assessment questionnaires, which is related to the self-assessment questions, the results for both years are shown in Fig. 2. The left chart shows the results for the first year while the second is related to the second year. On each of the charts, the results before and after the activity are shown. In this case, the magnitude represented by the chart is the average value obtained, following the Likert scale (from 0 to 4), by the students on each of the skills (outer values). It is worth noting that, in both years, the polygon related to the skills before the activity is contained in the polygon obtained with the final results. This fact shows that the students have improved their knowledge related to all the technologies considered. The results for both years are quite similar, but the improvement achieved in the second year was more obvious than the one obtained in the first.

To sum up, we can obtain some general conclusions related to the obtained results for both years:

- The students improved their general knowledge about ML and DS tasks.
- The individual results on each of the topics or technologies have also been improved and the students are aware of the improvement, as was shown by the self-assessment results.
- The level of interest and participation on the second year was higher than in the first one.

4.2 Results of the teaching innovation experience

The results obtained from this teaching innovation experience are as follows:

- The students were made aware of the importance and interest that the professional profile of Data Scientist has nowadays, as well as the growing demand that exists around this figure, becoming a successful professional career for the students of the BSc in Computer Engineering.
- The students were introduced to the various tasks performed by the Data Scientist, including knowledge of mathematical and statistical notions of the models used.
- The students were introduced to a series of professional and current technologies for modelling real problems using ML techniques, such as *python* and the *scikit-learn* library.
- It ensured that students were able to contribute and receive new practical knowledge from the large community on the *Kaggle* platform, made up of well-established researchers and Data Scientists from a wide range of fields.
- The use of ICT in the classroom was encouraged and good participation in the workshops, which increased in the second year, was achieved.
- Group work was promoted by encouraging cooperation between participants in the same group and healthy competition between groups, so that the feedback and relationships between them culminated in improved knowledge for all.
- The students' self-confidence to get involved in real problems where they can apply their theoretical and practical knowledge to provide a tangible solution was increased.
- Carrying out the same type of workshop in two different years helped reinforce the robustness of the results achieved with them, given that, in both years, the results were similar and in both cases were quite positive.

5 Conclusions

After analysing the results, we believe that this kind of TIPs substantially improve the student's ability to face a real work situation after finishing their degree. In this sense, we worked with the Data Scientist profile, one of the most demanded profiles in today's society, having a great impact worldwide. The projects developed across two different years provided students with the motivation and capacity for continuous learning and recycling of knowledge that graduates in Computer Engineering must undergo.

Undoubtedly, the degree of participation and interest of the students in both years was very high, especially in the second year, where a higher number of students took part in the workshops. They spent a lot of time working on the activities that were prepared for these projects due to the difficulty and the study of a novel field that has not been approached by them so far. As a result of these experiences in *Kaggle*, several students became interested in carrying out Final Degree Projects related to research, ML and DS.

It is worth noting that the effort made by the organiser team that worked on the TIPs was notable. It was necessary to adapt all the workshop material and the proposed work for the activity to the level of the students. Nevertheless, the level of satisfaction achieved made it enormously worthwhile. In the future, new editions are expected to be held to increase the knowledge of this job profile, which is currently in demand and growing during the last years.

Acknowledgements The two Teaching Innovation Projects have been funded by the University of Córdoba with references 2017-1-5008 and 2018-1-5015. This work has also been partially subsidised by the "Agencia Española de Investigación (España)" (grant reference: PID2020-115454GB-C22 / AEI / 10.13039 / 501100011033); the "Consejería de Salud y Familia (Junta de Andalucía)" (grant reference: PS-2020-780); and the "Consejería de Transformación Económica, Industria, Conocimiento y Universidades (Junta de Andalucía) y Programa Operativo FEDER 2014-2020" (grant references: UCO-1261651 and PY20_00074). David Guijo-Rubio's teaching was funded by the University of Córdoba through grants to Public Universities for the requalification of the Spanish university system from the Ministry of Universities funded by the European Union - NextGenerationEU (Ref. UCOR01MS). The teaching of Víctor M. Vargas was funded by the "Programa Predoctoral de Formación al Profesorado Universitario (FPU)" of the Ministry of Science, Innovation and Universities (Ref. FPU18/00358). The teaching of Antonio M. Gómez-Orellana was funded by "Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía" (Ref. PREDOC-00489).

References

1. van der Aalst, W.: Data Science in Action, pp. 3–23. Springer Berlin Heidelberg, Berlin, Heidelberg (2016)
2. Alpaydin, E.: Machine learning. MIT Press (2021)
3. Brunner, R.J., Kim, E.J.: Teaching data science. Procedia Computer Science **80**, 1947–1956 (2016)
4. Brynjolfsson, E., Mitchell, T., Rock, D.: What can machines learn, and what does it mean for occupations and the economy? In: AEA Papers and Proceedings, vol. 108, pp. 43–47 (2018)
5. Bureau, E.T.: 11.5 mn job openings by 2026, sky-high salaries: Why data science is booming. The Economic Times (2020). URL https://bit.ly/3tsDVwZ
6. Caponetto, I., Earp, J., Ott, M.: Gamification and education: A literature review. In: European Conference on Games Based Learning, vol. 1, p. 50. Academic Conferences International Limited (2014)
7. Chesani, F., Galassi, A., Mello, P., Trisolini, G.: A game-based competition as instrument for teaching artificial intelligence. In: Conference of the Italian Association for Artificial Intelligence, pp. 72–84. Springer (2017)

8. Chow, W.: A pedagogy that uses a kaggle competition for teaching machine learning: an experience sharing. In: 2019 IEEE International Conference on Engineering, Technology and Education (TALE), pp. 1–5. IEEE (2019)
9. Columbus, L.: Linkedin's fastest growing jobs today are in data science and machine learning. Retrieved April **10**, 2018 (2017)
10. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: Defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek '11, p. 9–15. Association for Computing Machinery, New York, NY, USA (2011)
11. Durán-Rosal, A., Fernández, J., Gutiérrez, P., Hervás-Martínez, C.: Detection and prediction of segments containing extreme significant wave heights. Ocean Engineering **142**, 268–279 (2017)
12. Garcia-Algarra, J.: Introductory machine learning for non stem students. In: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, pp. 7–10. PMLR (2021)
13. Guijo-Rubio, D., Casanova-Mateo, C., Sanz-Justo, J., Gutierrez, P., Cornejo-Bueno, S., Hervás, C., Salcedo-Sanz, S.: Ordinal regression algorithms for the analysis of convective situations over madrid-barajas airport. Atmospheric Research **236**, 104798 (2020)
14. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. ACM SIGKDD explorations newsletter **11**(1), 10–18 (2009)
15. Hicks, S.C., Irizarry, R.A.: A guide to teaching data science. The American Statistician **72**(4), 382–391 (2018)
16. Jackson, A.E.: The 50 best jobs in america for 2018. Glassdoor (2018). URL https://bit.ly/3vN5EKL
17. Kross, S., Guo, P.J.: Practitioners teaching data science in industry and academia: Expectations, workflows, and challenges. In: Proceedings of the 2019 CHI conference on human factors in computing systems, pp. 1–14 (2019)
18. Lee, J.J., Hammer, J.: Gamification in education: What, how, why bother? Academic exchange quarterly **15**(2), 146 (2011)
19. Legaki, Z., Hamari, J.: Gamification in statistics education: A literature review. In: GamiFIN Conference 2020, CEUR workshop proceedings, pp. 41–51 (2020)
20. Miller, S., Hughes, D.: The quant crunch: How the demand for data science skills is disrupting the job market. Burning Glass Technologies (2017)
21. Milosz, M., Milosz, E.: Gamification in engineering education–a preliminary literature review. In: 2020 IEEE Global Engineering Education Conference (EDUCON), pp. 1975–1979. IEEE (2020)
22. National Academies of Sciences, Engineering, and Medicine and others: Data science for undergraduates: Opportunities and options. National Academies Press (2018)
23. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
24. Polak, J., Cook, D.: A study on student performance, engagement, and experience with kaggle inclass data challenges. Journal of Statistics and Data Science Education **29**(1), 63–70 (2021)
25. Ramamurthy, B.: A practical and sustainable model for learning and teaching data science. In: Proceedings of the 47th ACM Technical Symposium on Computing Science Education, pp. 169–174 (2016)
26. Sailer, M., Homner, L.: The gamification of learning: A meta-analysis. Educational Psychology Review **32**(1), 77–112 (2020)
27. Stack-Overflow, I.: Stack overflow developer survey 2021 (2021). URL https://bit.ly/3y0KhbB