laSalle
UNIVERSITAT RAMON LLULL

Empowering Curiosity: Bridging Academic Learning and Impactful Research in the Master in Big Data

Ester Vidaña-Vila, Víctor Aguado
HER – Human Environment Research
La Salle Campus Barcelona – Ramon Llull University
Barcelona, Spain
ester.vidana@salle.url.edu, victor.aguado@students.salle.url.edu

Joan Navarro
Smart Society
La Salle Campus Barcelona – Ramon Llull University
Barcelona, Spain
joan.navarro@salle.url.edu

***Abstract*— Stimulating students' curiosity toward science is crucial for their professional development and growth. However, as industry is more appealing than academia in the eyes of most students, seducing them with science can be challenging. This situation is further emphasized in masters' programs where graduate students pursue a specific specialization that enables them to open certain opportunities in industry and often neglect academia. Through this narrative, this work aims to provide insights into the broader implications of linking classroom education with tangible research outcomes in the context of a master's degree (i.e., Master in Big Data). Specifically, this paper describes the process that has enabled a student to seamlessly transition from a project assignment in an individual subject (i.e., Data Mining) to an accepted paper in an international conference. Although some students' final thesis are sometimes refined and extended to be submitted to scientific conferences, this work focuses on the early stages of such process. Hence, it sheds light on the potential of educational innovations to empower students, attract talent to scientific inquiry, and bridge the gap between academic learning and impactful research. The presented case study serves as an inspirational model for educators and institutions seeking to enhance the research culture within academic programs, particularly in the dynamic realm of Big Data.**

Keywords—research, data mining, acoustic event detection

I. INTRODUCTION

Universities play a crucial role when linking the knowledge generation (i.e., scientific research) with its transmission (i.e., education). Occasionally, this ends up with a beautiful loop backwards in which students exposed to the education process contribute to the knowledge generation process. For instance, there are a few final theses [1-4] that become the seed of a journal or conference paper, which subsequently exposes the student to the scientific knowledge generation process. Usually, this happens in a context where students are given the appropriate time, opportunities, and motivation from the faculty staff. Establishing this fruitful context in the hectic discourse of specialization courses with very condensed syllabi is challenging. Therefore, this work reports the experience of setting up this context for stimulating scientific curiosity in students enrolled in a master's degree.

The selected master's degree has been the Master in Big Data Engineering from La Salle Campus Barcelona. This program aims to train students in the usage and development of modern technologies to acquire, store, process, and analyse large amounts of data (i.e., Big Data). It is a one-year program of 60 ECTS credits. In this industry-oriented program, where innovation, technology, and practical application intertwine, students are eager to learn everything that may help them developing the required hard and soft skills to meet the needs of an ever-growing industry market. Although these needs often require a considerable amount of research and development, students are usually blinded by the appealing job positions that industry offers, which often drives them to neglect the academic market and opportunities. Therefore, we believe that the Master in Big Data program stands as a challenging crucible for cultivating curious and scientific minds.

Creating the necessary context to stimulate students' scientific mindset in such environment has been achieved by (1) planting the seed of curiosity in scientific methodology in the Data Mining subject that has nine 3-hour sessions (i.e., 6 ECTS) and (2) gathering its results in the Final Thesis subject. The result of this experience has been materialized in an accepted paper that was presented in the Detection and Classification of Acoustic Scenes and Events (DCASE) 2023 international conference [5].

II. USE CASE

To ignite the scientific curiosity of the students, the practical assignments of the Data Mining subject were carefully planned. First, the foundation of the practical assignments was laid by selecting a topic closely inspired by existing research within the Signal Processing line of the Research Group in Media Technologies. Hence, they were introduced to a real-world application of data mining techniques within the domain of acoustic signal processing. The deliberate choice of a research-oriented project aimed to expose students to the complexities of real-world challenges and foster a research-oriented mindset. Then, the final project of the subject consisted on developing a bird classifier using acoustic signal processing and machine learning and deep learning techniques. Even though there are many bird datasets available online, the students were told to use a dataset crafted by an alumnus in the context of his Master's Final Thesis [4]. The selection of this dataset was not arbitrary, it would enable to perceive part of the research carried out in university, and, moreover, that a master's student like them could also contribute to such research. While the presented paper was intended as a baseline, students were motivated to explore new techniques, aiming to improve classification metrics. This brought a fierce competition among them and their efforts towards surpassing the paper's classification outcomes.

After presenting the initial assignment, the seed of curiosity was sown. The next step was to propose an attractive Final Master's Thesis project to students that could compete with other industrial projects. The proposal consisted of participating in the DCASE 2023 challenge, an international competition that serves as a platform for researchers to advance the field of acoustic scene analysis and event detection. This challenge presents a set of standardized tasks related to the analysis of large and complex audio files (i.e., Big Data). These tasks are designed to mimic real-world challenges, ranging from identifying common sounds in urban environments to more specialized tasks like recognizing specific acoustic events. Participants in each task are then invited to present their approaches in the conference.

To dive deeper in the topic explored in the Data Mining subject and the possibility of participating in an impactful scientific conference, one student selected the proposed scientific project and was given the freedom to select any of the tasks offered in the challenge. Predictably, the student opted for Task 5 [6], focusing on few-shot learning applied to the bioacoustics domain. This resonated closely with the topics explored during the course, aligning seamlessly with the hands-on experiences already gained.

While the student demonstrated exceptional proficiency in algorithm development, timing constraints prevented their official participation in the contest. Nonetheless, the noteworthy results achieved were substantial enough to warrant the creation of a peer-reviewed scientific publication for the DCASE workshop [5], which was accepted for publication after a peer-reviewed process [6].

CONCLUSION

This work has presented a distinctive approach to fostering scientific curiosity within the Master in Big Data Engineering master's program that lasts one academic course. In this experience, it has been shown how scientific curiosity can be stimulated using a plot thread (rather than the typical one-shot strategies) from a multifaceted perspective. That is, (1) selecting appropriate topics and challenges for classroom-oriented projects and assignments to provide the necessary background and mindset, (2) offering an appealing methodology (i.e., competition with world-class researchers), and (3) opening the opportunity for a nice recognition (i.e., publication in a scientific international conference) for the conducted job. We believe that this process has enriched the student, the researchers, and the university.

REFERENCES

[1] Navarro, J., Vidaña-Vila, E., Alsina-Pagès, R. M., & Hervás, M. (2018). Real-time distributed architecture for remote acoustic elderly monitoring in residential-scale ambient assisted living scenarios. *Sensors*, *18*(8).

[2] Blanch, J., Vidaña-Vila, E., & Alsina-Pagès, R. M. (2021). Analysis and Automatic Classification of the Soundscape of a Natural Park Close to the Airport of Barcelona during the 2021 Lockdown Period. *Engineering Proceedings*, *10*(1), 41.

[3] Vallés, V., & Vidaña-Vila, E. (2023). EduSign: Real-Time Application for Spanish Sign Language Recognition. In *Artificial Intelligence Research and Development* (pp. 132-133). IOS Press.

[4] Gómez-Gómez, J., Vidaña-Vila, E., & Sevillano, X. (2023). Western Mediterranean Wetland Birds dataset: A new annotated dataset for acoustic bird species classification. *Ecological Informatics*, *75*, 102014.

[5] Aguado, V., Navarro, J., & Vidana-Vila, E. Learning in the wild: bioacoustics few shot learning without using a training set. Proceedings of the 8th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE 2023), September 2023.

[6] Nolasco, I., Singh, S., Morfi, V., Lostanlen, V., Strandburg-Peshkin, A., Vidaña-Vila, E., ... & Stowell, D. (2023). Learning to detect an animal sound from five examples. *Ecological Informatics*, *77*, 1022