

## Abstract

The number of citizen science (CS) projects has grown significantly in recent years, owing to technological advancements. One important aspect of ensuring the success of a CS project is to consider and address the challenges in this field. Two of the main challenges in CS projects are sustaining participation and improving the quality of contributed data. Despite the studies that have been conducted to address these two challenges, there is still a need for new approaches, one of which is the use of artificial intelligence (AI) and machine learning (ML) in CS projects.

Therefore, the objective of this thesis was to investigate the integration of ML and CS, as well as the role of this integration in addressing CS challenges. A comprehensive review conducted in this study of motivational factors in CS projects indicated that *interest in learning about science* and *receiving feedback* were strong motivations among participants in the majority of CS projects. Typically, experts verify the data and provide feedback to participants. However, due to large amounts of data, this manual data verification can be time-consuming. Thus, in this research, it was investigated how the integration of ML and CS can, on the one hand, automate and speed up the data validation process, and on the other hand, increase participant engagement and sustain participation by providing real-time informative feedback.

To that end, a biodiversity CS project was implemented with the goal of collecting and automatically validating observations as well as providing participants with real-time feedback. ML algorithms were trained to model species distribution using environmental variables (e.g., land cover) and species data from an existing CS project, and then to validate a new contributed observation based on the likelihood of observing a species in a specific location. Furthermore, volunteers were given real-time feedback on the likelihood of observing a species in a particular location, as well as species habitat characteristics. Moreover, a user experiment was conducted, and the results indicated that participants with a higher number of contributions found the real-time feedback to be more useful in learning about biodiversity and stated that it increased their motivation to contribute to the project. Besides that, as a result of automatic data validation, only 10% of observations were flagged for expert verification, resulting in a faster validation process and improved data quality by combining human and machine power. Finally, based on the findings of the experiments and the discussions that followed, we made some recommendations for CS practitioners to consider before designing a new project or improving an existing one.

The future objective of this research is to focus more on the challenges of ML and CS integration, and to investigate how this integration can be applied in other CS fields besides biodiversity.