

LEVERAGING NATURAL LANGUAGE PROCESSING FOR COMPREHENSIVE STUDIES OF SCIENCE STUDENT PROJECTS

James D. Cleaver^a, Laura K. McKemmish^a, and Sara H. Kyne^a

Presenting Author: James D. Cleaver (j.cleaver@student.unsw.edu.au)

^aSchool of Chemistry, Faculty of Science, University of New South Wales, Sydney NSW 2052, Australia

KEYWORDS: artificial intelligence in education, equity and inclusion, natural language processing

Student research projects are a crucial part of the Australian and New South Wales (NSW) High School Curriculum. In NSW, the extension science course offered for the Higher School Certificate is an example of an extensive project performed by students. The objective of the course is to provide students the opportunity to authentically apply scientific research skills. Extension science and related courses for high school students are commonly assessed through scientific reports submitted as a final summative assessment (Science Extension | NSW Education Standards, n.d.). This gives rise to large volumes of disparate data which can potentially be analysed for insights to improve science teaching and learning. Understanding these insights are especially important for priority groups to increase accessibility and equity and reduce academic attainment gaps in science.

Previous research analysing student projects has been limited to studying small numbers of projects, due to the availability of data and the time taken for manual data analysis. This also limits analyses to single diversity variables, such as ethnicity (Carlone & Johnson, 2007). There is an opportunity to be realised in the data from student projects that may inform how teachers can better cater for the needs of students in various priority groups moving forward.

This study outlines a method to address this research gap, by employing artificial intelligence (AI) capabilities, particularly natural language processing (NLP) techniques, to examine large sets of science high school students' final project reports such as those retained by student science fairs. A range of AI techniques have been evaluated to enable us to process and analyse sizable datasets to explore the rich information they contain. NLP techniques have been developed to classify and analyse projects along various dimensions, such as the alignment with the Field of Research (FoR) codes, the research themes. The dimensions identified will then be analysed and correlated with demographics relating to priority groups.

These methods are informing the development of a reliable and repeatable AI-powered framework to analyse research themes, amongst other variables contained within science students' final project reports. The goal of this framework is to inform the learning design of science projects to increase accessibility, student engagement and inclusion.

REFERENCES

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching*, 44(8), 1187–1218. <https://doi.org/10.1002/tea.20237>
Science Extension | NSW Education Standards. (n.d.). Retrieved 22 May 2023, from <https://educationstandards.nsw.edu.au/wps/portal/nesa/11-12/stage-6-learning-areas/stage-6-science/science-extension-syllabus>

Proceedings of the Australian Conference on Science and Mathematics Education, The University of Tasmania, 30 August – 1 September 2023, page 23, ISSN 2653-0481.