Undergraduate Research position in Physics: Using machine learning to analyze open-ended survey responses

**Job Description**
Lab courses are an integral part of physics curricula. Developing an understanding of measurement uncertainty is a common learning goal of lab courses. As physics education researchers at CU Boulder, we have been using a survey, the Physics Measurement Questionnaire (PMQ) [1], to measure students’ understanding of measurement uncertainty in order to understand and improve students’ experiences in lab courses. However, responses to the PMQ are open-ended and require significant time and effort to analyze. In collaboration with researchers at Michigan State University (MSU) and the University of Oslo (UiO), we are interested in exploring the use of artificial intelligence techniques in the field of natural language processing (NLP) to analyze PMQ responses.

We are offering an undergraduate research position to conduct preliminary investigations into using NLP to analyze PMQ responses. The student would start in the Spring 2019 semester at CU Boulder, working with researchers there to gain familiarity with the project goals, the lab courses being studied, and the PMQ itself. That work would be supported through independent study credit, or possibly financial compensation depending on funding availability. The student would then travel to Oslo, Norway for the Summer of 2019 to work with researchers there to learn and implement NLP techniques on PMQ data. Funding will be provided to support travel to Oslo, a summer stipend, and housing expenses there. While not guaranteed, continuing this research at CU after Summer 2019 is a strong possibility and could likely lead to an honors thesis at CU.

This project will be co-advised by Ben Pollard and Heather Lewandowski at CU Boulder, John Aiken at UiO, and Danny Caballero at MSU.

**Job Requirements**
Applicants should be current undergraduate students at CU Boulder with an interest in machine learning, natural language processing, and/or science education. They must be available as a student researcher in Summer 2019 and the 2019-2020 academic year. Applicants must have completed or be currently enrolled in PHYS 1140. Prior programming experience is a must for this position, ideally with concurrent enrollment in a computer science course beyond the intro level in the Spring 2018 semester.

**Skills learned through this research project**

1. Python programming using pandas, scipy, numpy, and NLP libraries
2. Statistical and machine learning techniques related to classifying topics within written English documents
3. Scientific writing and presenting scientific results
4. Potential exposure to technologies like SQL, R, and other common data science tools

**Job Application**
To apply for the position, please send the following materials to Heather Lewandowski (lewandoh@colorado.edu) by 12 October 2018.

1) Curriculum Vitae
2) Unofficial transcript
3) Statement of your interest in the position including addressing the following points.
   a. Why do you want to participate in research and this project in particular?
   b. Describe your previous experiences with computer programming (Note: we will expect you to have some previous coding experience (C/C++, FORTRAN, or Python)).
   c. How many hours would you like to work on the project during the Spring 2019 semester? Would you prefer independent study credit or to be paid?
   d. What coursework have you taken, and will you take over the next year that is related to programming, machine learning, and/or natural language processing?

References