



CYBERSECURITY RESEARCH: CLICKBAIT OR NO-CLICKBAIT

Miral Ahmad¹, Helena Mentis², Nora McDonald³

¹Department of Information Systems, University of Maryland, Baltimore County 1, 1000 Hilltop Cir, Baltimore, MD 21250

I have participated in the Louis Stokes Alliances for Minority Participation. I had partnered up with Dr. Helena Mentis and Mrs. Nora McDonald in the department of Information Systems and had my research had focused on the Cybersecurity field. I have been interested in research and want to obtain a Ph.D. in Information Systems because networking is my passion and technology is rapidly growing.

What my task for the research was that I had to utilize the programming language Python to determine if the output of the code had made all of the technical articles clickbait or no-clickbait. Specifically for the code, what it does is it parses through the instances.jsonl file and converts the parameters such as post stamp, to an object. The next step was that I parsed the truth.jsonl file that contains the truth value such as truthJudgements and I converted those parameters to an object. Both of those objects have gotten appended to a list and then it was ready for Natural Language Processing.

Finally, what I did was I split the data with a 70-30 ratio, which is the recommended data split between testing and training. So, the 70% is the training data my model trains on and tests against the 30% data to verify its accuracy. I vectorized the target paragraphs to represent integers instead of words. The process of vectorization simply means converting words to numbers. There are a variety of implementations, but the one I used is a TFID where it converts the most commonly used English words and gives it a score. So, words that are more repetitive like: the, a, is, etc are given less weight. Therefore, better identifying a clickbait article. In conclusion, I have enjoyed working with my mentors and have learned more about technical research.