Capstone Milestone 1

Rosalind Kidwell, Ilana Seidl, and Chase Stockwell

1. For our project, we will study gender bias in financial literacy resources, especially cryptocurrency. We will show how Artificial Intelligence contributes to cyclical gender bias. Our finalized product will be an algorithm that assesses a resource on its gender bias.

We will receive our information from an amalgamation of resources. We will use online database sets to compile data, professors from both the Finance, Linguistics, and Gender & Sexuality major departments, and information sets that Dr. Mettei, our mentor for this project, recommends to us.

1. Gender-biased writing is prevalent in nearly any space in cryptocurrency, as the industry is heavily male-dominated. Anyone currently writing or developing work on cryptocurrency would be our target readers, so that they can understand the prevalence of this problem, ways they can address it, and better understand why women are not investing at the rate that men do in cryptocurrencies.

Most other studies about women’s financial literacy do not discuss cryptocurrency, as this is a relatively new area of the financial world. Studies show that economic textbooks greatly underrepresent women in their given examples[[1]](#footnote-0) and that popular online forums about economics are quite sexist.[[2]](#footnote-1) These studies affirm our understanding that financial resources can be quite gender-biased, but none of them deeply look into the cryptocurrency space.

In relation to algorithms that analyze gender bias, the most prevalent study we’ve found is “Identifying and Reducing Gender Bias in Word-Level Language Models[[3]](#footnote-2).” In this study, a metric is proposed to measure gender bias in a text corpus. The researchers calculated probabilities of certain words being associated with one gender over the other. A text is considered to be biased if there is a skew of words co-occurring with one gender over the other. This is a similar method that we can take up when analyzing texts for our study.

1. Our group seeks to highlight the devastatingly low numbers of women participating in cryptocurrency. We plan on developing a measurement tool to help indicate gender-bias in the space. To do this, we plan on gathering data and building a system using Natural Language Processing (NLP) to process and understand biases in language used around the emerging cryptocurrency field.Through NLP, we plan to highlight gender-bias on a fundamental, meaningful level. We hope that this will enable us to develop an accurate measurement of bias in cryptocurrency and beyond.

Our approach involves several phases. Firstly, we are gathering and cleaning data to streamline our findings into meaningful and relevant datasets. Following this, we plan to explore several tools including NLP and sentiment analysis. We will be using tools such as Docker, Jupyter Notebooks, and programming in Python. With this, we are also aware that we are going to utilize new skills, languages, and techniques that we are currently not aware of.

Through NLP, we plan to build an algorithm that will read, understand, and derive meaning from the various data sets we collect. Potential issues include developing a balanced algorithm that assesses gender-bias appropriately. In addition, obtaining enough relevant data may pose an issue. Ideally, we will build a model with low-resources in mind.

More specifically, through sentiment analysis, we plan to classify information as biased or unbiased. Using this will aid us in understanding gender-bias in the cryptocurrency space. Potential issues include using a classification system that is either too strict or too loose and missing valuable information.

Our approach takes a unique stance on quantifying and understanding the world of cryptocurrency. At the moment, there is minimal published work on bias in cryptocurrency and the potential effects of excluding populations from this new technology. With this, we plan to utilize existing Word-Level Language Models and past research on identifying gender bias in specific fields.

1. Our biggest concern for feasibility will be gathering sufficient data to derive any formulas on bias. To address this, we will use a wide variety of sources for data including tweets, forums, news articles, scholarly articles, and informative articles about how to use cryptocurrencies. We also need to establish clear objective measures of what makes a text biased. For this, we plan on studying about language differences by gender. We’d also like to perhaps meet with some linguistics professors and gender and sexuality professors to find clear markers of a biased text, learning what makes a text gender inclusive. As discussed with Dr. Mattei, our goal by the end of the semester is to have completed running a large data set through the NLP pipeline in order to gain a better understanding of NLP, before applying it to our specific data set. We have been meeting Wednesday afternoons with Dr. Mattei and will continue that schedule, meeting as needed throughout the semester, ensuring that we meet as milestones approach.
2. The project is divided into three components. Chase will research regarding crpytocurrency’s history, data valuation, and specific examples of gender discriminatory behavior in cryptocurrency. Rosalind will research on NLP methods and topic modeling to create the technical structure of the project. Finally, Ilana will research on what elements contribute to gender bias in a resource, what lexicon is gender-biased vs. gender neutral, and methods to eradicate gender bias.

This project will involve significant work outside of class. Chase plans on spending four and a half to five hours a week gathering data sets regarding cryptocurrency. Rosalind plans on spending four to five hours a week learning about NLP methods to create the ratings scale. She’ll also meet with professors from the linguistics department to gain more background on linguistic differences based upon gender. Ilana plans to spend three and a half to five hours a week gathering public data sets as well as meeting with gender and sexuality associates at Tulane.

1. Flaherty, C. (2018, January 19). Women are underrepresented in economics textbooks, says a new analysis, with implications for the field's gender imbalance. Retrieved from https://www.insidehighered.com/news/2018/01/19/women-are-underrepresented-economics-textbooks-says-new-analysis-implications-fields [↑](#footnote-ref-0)
2. Wu, A. H. (2017). Gender Stereotyping in Academia: Evidence from Economics Job Market Rumors Forum. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3051462 [↑](#footnote-ref-1)
3. Bordia, S., & Bowman, S. R. (2019). Identifying and Reducing Gender Bias in Word-Level Language Models. *Proceedings of the 2019 Conference of the North*. doi: 10.18653/v1/n19-3002 [↑](#footnote-ref-2)