Name: Jerin Easo Thomas
Semester: Spring 2023

Title: Spotting the unseen: Identifying Sexism in Twitter Movements using Machine Learning and Deep Learning approach. (583 words)

Research Question:

1. How reliable and efficient are machine learning methods in spotting sexism in Twitter movements?
2. How do various social media movements differ in terms of sexism types and source intention?

Background:

Social media remarks that spread damaging preconceptions about women or gender-based discrimination are referred to as sexism comments. The language used in these remarks can range from more covert forms of discrimination, such as microaggressions, to more overt language that is intended to humiliate, objectify, or frighten women. To create safe and welcoming environments for all users, stop the spread of harmful and discriminatory attitudes, lessen the psychological effects on people, and gain insights into patterns and trends in gender-based discrimination to advance gender equality, it is critical to analyze and identify sexism and its source intension on social media tweets.

Angel Felipe (2022) talks about how difficult it is to recognize and categorize sexist content in social media posts, especially given things like sarcasm and the many different shapes that sexism may take. The performance of various transformer topologies, including BERT, RoBERTa, Electra, and GPT2, for sexism detection and classification tasks in English and Spanish is evaluated by the author along with an approach for solving this issue [1]. Adding on *Shimi Gersome (2022)* talks about how deep learning models can be used to spot sexism in online forums. The author describes a method that uses a convolutional neural network (CNN) and a pre-trained word embedding model to categorize tweets as sexist or non-sexist. The study concludes that deep learning models can be useful for detecting sexism in social media and that additional research is required to enhance the system's performance for detecting various forms of sexism and resolving the problem of false positives [2].

Data:

The three distinct Twitter moments #MeToo, #8M, and #Time'sUp will provide the data for this study. The time period for the tweets would be from the movement's inception until the present; for instance, the #MeToo movement mostly began in 2017; hence, the data would be collected from October 2017 until the present. The tweets based on the case-sensitive search parameters "#MeToo" "#8M" and "#Time'sUp" would be fetched using a social networking service scrapper (SNSCRAPE) written in Python. 500 tweets from each of these twitter moments would be pulled to ensure equality, for a total of 1500 tweets. Along with the tweet comment field, other variables include the tweet URL, tweet date, like and view count, and username.

Method:

The tweets would first be divided into two categories: "Sexist" and "Not Sexist." Once the tweets have been determined to be sexist, the next task aims to classify them into "Direct" which is the intention to write a message that is sexist in itself or incites others to do the same, "Reported" which is the intention to report and share a sexist situation experienced by women and "Judgmental" the intention was to judge and condemn the opposite gender. To perform these classification Robustly Optimized BERT Pre-training Approach (RoBERTa), a machine learning model, and Neural Network, a deep learning model in Python, would be used. At the beginning, the training data for these models would come from the website called *EXIST* [3], allowing them to be built using sexist tweets, and the test data for these models would be the tweets data obtained through SNSCRAPE. To deduce the best differences between moments in terms of sexism types and intention, the model that produces the best result and has the highest accuracy would be chosen for further analyses and visualization.

References:

[1] Angel Felipe Magnossão de Paula and Roberto Fray da Silva. Detection and Classification of Sexism on Social Media Using Multiple Languages, Transformers, and Ensemble Models. Retrieved September 2022, from https://ceur-ws.org/Vol-3202/exist-paper2.pdf

[2] Shimi Gersome and Jerin Mahibha. Sexism Identification In Social Media Using Deep Learning Models. Retrieved September 2022, from https://www.researchgate.net/publication/364947922_Sexism_Identification_In_Social_Media_Using_Deep_Learning_Models

[3] Francisco Rodriguez-Sanchez, Jorge Carrillo-de-Albornoz, and Laura Plaza. Automatic Classification of Sexism in Social Networks: An Empirical Study on Twitter Data. Retrieved December 2020, from https://ieeexplore.ieee.org/document/9281090

[4] EXIST: sEXism Identication in Social Networks. Retrieved from http://nlp.uned.es/exist2023/