

Proposal

Magdaleno, Alejandro
`amagale@student.kennesaw.edu`

Lyles-Woods, Quin'darius
`qlyleswo@student.kennesaw.edu`

September 24, 2021

Proposed Study

For the topic of study we have decided to venture into the arena of political language. More specifically we are looking to understand more about the voice of prose and seeing if that can be algorithmically detected with Natural Language Processing. We will build a classification model to determine what news outlet a given body of text has originated from.

Goals

The goals of the proposed study to build a classification model to determine where bodies of text originate from will be first and for most to be able to, of course, detect where the bodies of text originated from. An auxiliary goal will be to tell which of the author has written the given body of text. Another auxiliary goal will be the level of accuracy the model will have. Aiming to be better the 70% to begin with and gradually increasing the threshold. If we pass the 95% mark we will considered the first auxiliary goal mentioned which is to classify the authors voice in a given piece.

Relevance

The relevance of the proposed study of algorithmically detecting the origin of new articles can be linked to the analysis of bias and unearthing it within the large repertoire that we are given with our selection of where to find and digest news of the world around us. If we can find out the bias this should be the first step and shifting through to the truth in these sometimes lengthy articles that share more about the authors voice than the subject matter at hand. Doing this will save a great deal for anyone that just wants the facts fast and will allow those individuals to regain parts of their lives that would have been ultimately lost other wise.

Data

Models and Algorithms