

While more and more news providers are recognizing the importance of news relating specifically to women, some still lack clear sections dedicated to this kind of content. This oversight usually makes it difficult to search for articles that deal with women's issues such as reproductive rights, economic justice, and ending violence against women, especially when searching for the most recent and relevant news. While article-based news organizations are slightly better at this, podcast providers are still lagging behind. One such source that is missing this collection of similar content relating to women is National Public Radio (NPR). The proposed application would aggregate all NPR podcast episodes and news articles relevant to women and present them in a feed with the most recent being first.

The proposed project would take advantage of NPR's application program interface (API) to retrieve podcasts and news stories from their large archive that dates back to 1995. The JSON that will be returned from the API request will then be parsed to find articles and podcasts that related to women. The Multinomial Naive Bayes model will be used to estimate the probability of a particular word taking frequency into account. The Naive Bayes Classifier will be used over other text analysis algorithms, such as Max Entropy or Support Vector Machines, because it requires a relatively small amount of training data, which in the context of this 8-month project makes its training time significantly smaller and more desirable. It is also overall less computationally intensive than other algorithms both in memory and CPU.

**Intellectual Merit:** The NPR archive is relatively large at around 250,000 stories. This could pose a problem where it is too big to be processed by an algorithm. While the Naive Bayes Classifier has its own feature extractor, which is relatively simple, this might not suffice for a dataset this large. The proposed project would develop its own feature extractor to significantly reduce the set of features that would still contain the relevant information from the input data. Therefore the task can still be completed by using this new condensed representation instead of the entire original dataset from the NPR archive.

**Broader Impacts:** The resulting application can be used by a wide variety of people who have an interest in women's issues. It could prove to be particularly helpful for students taking women's studies classes so they can easily keep up to date with the news. And since the application would include podcasts, it would make it easy to consume news and gain more knowledge in situations where reading a newspaper, magazine, or online article is not an option, like when driving a car. Another group of people this proposed application could help is people with dyslexia who want to stay informed on women's issues. Since reading is sometimes very challenging for people with dyslexia, the option to listen to podcasts would be ideal for them by lifting the burden of the extra mental concentration it takes to read through long articles.