IS507 - Final Project Literature Review Draft

Section BC - Group 16: Online News Popularity

(Omar, 2007) [1] seeks to observe how the transition from conventional news to online news has occurred and which factors have contributed to online news becoming popular. Immediacy is a factor that can be considered one which has been shown to influence both instant gratifications for the reader and allow readers to have faster access to the latest issues. This also means a certain news category will spread faster than others, which can be determined by understanding the polarity. (Shirsat et al., 2017) [2] successfully estimates the polarity of words (positive, negative, neutral). Using text mining and Term Document Matrix, sentiment analysis is performed at document and sentence levels so that if the sentence is positive, the document is positive, and if the document is negative, the sentence is negative. The articles were divided into genre groups based on their sentiment scores and classified as positive, negative, or neutral. Unlike the previous two journal articles (Deshpande, 2017) [3] identifies the optimal model to predict the popularity of online news by building an automated decision support system without considering the effect of sentiments or polarity of the article content. (Omar, 2007) [1] helps us understand the hypothesis testing that could be used. The application of the polarity of words and its analysis can be determined by (Shirsat et al., 2017) [2], and the

various machine learning methods that can be used to predict the accuracy of our model can be determined by (Deshpande, 2017) [3] which can further be improved.

Firstly, before performing any research techniques on the data, we intend to find out the most critical features that can be used to perform the analysis. The study done by (Deshpande, 2017) [2] related to the Linear Discriminant Analysis technique for performing dimensionality reduction will be useful here. Next, we will be working on predicting the number of shares/popularity of online news by using appropriate machine learning methods. We will work on finding similar articles based on the polarity of the words in those articles, which may be related to positive, negative, or neutral sentiments, as done by (Shirsat et al., 2017) [2]. A hypothesis that we will work on is whether the type of data channel (entertainment, lifestyle, etc.) influences the number of shares of the news articles. Another hypothesis would be to check if there is any association between when the online article is read and its number of shares.

We hypothesize that papers with negative sentiments will be more popular than those with positive sentiments, as an article dealing with negative news that can bring about harm is more likely to be shared amongst the masses. The distribution among data channels will be more or less uniform since we do not have any information about the types of readers and their preferences. The number of unique words, the number of images, the length of the article, and the sentiment (both positive and negative) can be some of the key features in deciding the popularity of an article. Because people are more relaxed on weekends, an article published on a weekend will be more popular than one

published on a weekday. An article with a lower number of words will be more popular than an article with a huge amount of content since it will give the information to the reader directly in a summarized format. We plan to learn the impact of negative and positive sentiments on news consumption and understand how it can affect the shareability of the news. This study can help us understand the type of content to be incorporated into online articles to ensure maximum outreach.

References:

[1]   Omar, B (2007). *The switch to online newspapers: Could immediacy be a factor?*

[2]   Shirsat, V. S., Jagdale, R. S., & Deshmukh, S. N. (2017). Document level sentiment analysis from news articles. *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. https://doi.org/10.1109/iccubea.2017.8463638

[3]   Deshpande, D. (2017). Prediction & Evaluation of online news popularity using Machine Intelligence. *2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. https://doi.org/10.1109/iccubea.2017.8463790