ALY6080: Signature Assignment Individual Project Proposal

Harnessing Social Media Data to Predict TV Ratings and Propose Successful Marketing Strategies

Rebecca Leu

Analytics, Northeastern University

ALY6110: Integrated Experiential Learning

Dr. Matt Goodwin

February 21, 2021

**Statement of Purpose**

Social media has become an essential part of our modern world. Although first catering to the younger generation of Americans, the news quickly spread of its benefits of connecting people old or young all over the world. People can share their experiences, lives and opinions all the time and often without limits. This has created an enormous amount of data about almost anything imaginable. Advertisers use basic understanding of a person’s online media presence to advertise to specific groups of people based off their age, occupation, political view, likes, dislikes, etc. The idea of this project is to harness the use of social media data to predict TV ratings as well as understanding best market strategies for advertisements.

There are two main goals of this project. The first is to use social media chatter to predict TV ratings. This would help to better determine TV ratings then current data provided by A.C Nelson’s use of sampling audiences. The second goal is to harness similar social media data to determine best avenues for each type of advertiser, during each type of show. This project would provide more or a dashboard to show different sectors of businesses how to get the most engagement out of their advertisements.

**Scope of the Project**

Although both parts of this project are using social media data to provide insight, the projects will need to be worked on separately. The first deliverable will be the TV ratings prediction model. This model will be built using prior shows social media chatter that have either done well, or not. The model will predict the outcome of the show based off the data. Once the model has been trained to the social media data and giving accurate results, it can be applied to new shows without rating data already known.

The second deliverable will be the dashboard that that helps advertisers determine how to shape their advertisements as well as when to air them to get the most engagement. The dashboard will be split into show types (drama, comedies, news, sports etc.) and then give engagement scores based off the type of products/services being advertised as well as additional analysis showing which type of advertisement moods preformed best and worst in the past. This will help media companies determine pricing for advertisers as well as promising engagement metrics to be expected by the advertiser to promote sales of the ad spaces.

**Background Research and Literature**

There has been some work or analysis already done in this area surrounding predicting tv ratings using online word of mouth, audience measurement using big data, as well as advertisement engagement around social media.

The first article went over utilizing online chatter to predict TV ratings. The paper investigates how TV ratings could be better predicted using online word of mouth. The media industry heavily relies of TV ratings as they use this for negotiating advertising fees and selecting programs to buy. There are many different models out there already to predict tv ratings, but none of them included online sources and most of them use conventional statistical techniques like regression (Yeh, 2015). The author in his experiment used reviews and TV ratings information from Yahoo! Movies and XYZ company to use as a predictor. This article relates to the business question of how to incorporate more social media, or outside online sources to benefit the media industry.

The paper describes in detail, the hybrid model used to complete this analysis. GRA, grey relational analysis is used in this study because it only requires a small sample size, does not use typical distributions, and doesn’t not involve independent factors. Yeh first incorporates the online word of mouth as a potential predictive variable and then, using GRA, selects the variables that are most important in predicting the ratings. After determining the most important predictor variables using GRA, they are used as inputs into an RST model. Ultimately, the model built by Yeh out preformed the usual predictor models and proved that online word of mouth can give more accurate predictions then a model without the use of this data (Yeh, 2015).

In addition to this, another article goes over the challenges of audience measurement using big data. This article describes how big data could and should be used in audience measurement. One of the main business for this is the market research company A.C. Nelson. For years they have been using the same basic way of gathering data by sampling audiences (Kelly, 2017). The author, JP Kelly, describes how big data should be used and how social media metrics play a role in these measurements as well. This article relates to the business question of how to incorporate more social media, or outside online sources to benefit the media industry.

Kelly describes that many media companies rely on audience measurement and tv ratings so much that it is hard to change the method of these ratings as it could mean millions of dollars lost if the new method does not work well. Kelly also points out the 6 provocations of big data and if the pros outweigh the cons. Kelly’s main points are how bigger data is not always better data, that limited access to big data can create digital divides and that automating research changes the definition of knowledge. His conclusion is that real time analytics promotes increased investment in certain genres, the scale of data may increase chances of perceiving meaningful connections between unrelated things, as well as giving larger companies the upper hand when being able to gather and use this data more than others (Kelly, 2017).

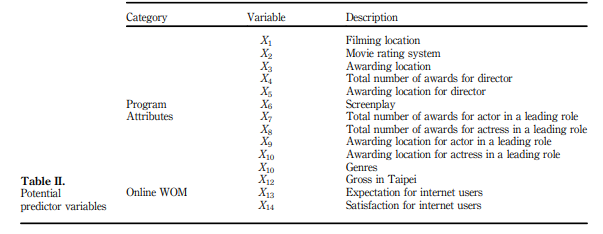
The third article is specifically about how social media activity interacts with TV advertising. This article relates to harnessing the use of social media data for understanding best market strategies for advertisements. The main function of this article is to discuss whether social TV is good or bad for advertisements. Social TV is known as watching a show while also engaging in social media chatter about it at the same time. Social TV shows are often viewed live and have a large audience discussing the events as they are unfolding on social media. Shows like the bachelor are a good example of this type of media. The main points being discussed are if advertisements during these types of shows are getting really engaged large audiences or if the audience is really engaged in the show while taking the commercial breaks to engage in social media about what just happened on the show. The writers analyzed over 9000 advertisement instances, 15 product categories on 84 prime time programs (Fossen & Schweidel, 2017).

There was a mixture of engagement, with some shows having more ad engagement then expected, and others having lower engagement than expected. There was a lot of differences in online engagement for types of ads with ads for movies and phones computers and tablets generating the most chatter and apparel, nonprofit and dental care ads generating the least amount of chatter. Some other interesting results of the analysis shows that longer ads and ads with hashtags created more online chatter. Additionally, the analysis saw that ads with funny moods and ones that mentioned price preformed best, while active, informational, and sexy advertisements resulted in a decrease in subsequent purchases on the retailer’s website (Fossen & Schweidel, 2017).

**Design, Data Collection Methods and Implementation**

The data to be used for this project would be twitter hashtag results, a shows or products Facebook or Instagram page comments, and advertisement comments and engagements. This data would have to be purchased or downloaded from these sites. Initially we would perform sentiment analysis’s and create graphs and tables to show correlations with TV shows and engagements as well as how advertisements are fairing with these programs.

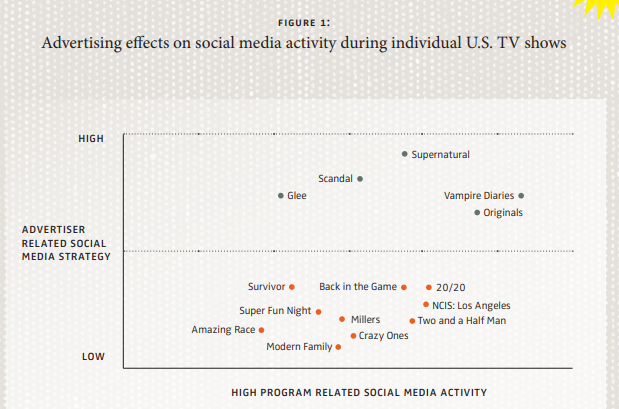
For the predictive model to determine TV ratings based off the social media data, I propose we use natural language processing of the social media data to be used as a predictor in a tv rating model. The tv rating model would also consist of other variables used to predict tv ratings with the addition of the NLP data as predictors. The natural language processing could be sentiment analysis or aspect mining. The proposal is to use a similar model as Yeh used in the research paper mention earlier. Shown is the variables he used as well as his online word of mouth variables.



(Yeh, 2015).

The type of model we use could be whatever is currently being used to predict TV ratings, with the addition of our social media data as optional variables.

For the interactive dashboard to be created for advertisement analysis, we would again use the same type of social media data with natural language processing applied to get a sense of the response online to the TV program and/or advertisements associated with it. This data could then be used to show advertisers which programs their advertisements might preform best with as well as what kind of mood of the advertisement works best for a particular audience. The figure below is an example of this analysis (Fossen & Schweidel, 2017).



**Conclusion**

Social media data can be utilized by analyzing more than just click rates and costs. By using natural language processing on social media chatter about a particular show or advertisement, we can predict TV ratings as well as propose successful marketing strategies for our own shows as well as to our advertisers. This data has a history of being relevant and useable in this type of analysis and is proven to be beneficial in these instances.