Joshua Mazher

DS 4002

December 9th, 2024

Case Study: Using Sentiment Analysis to Create a Model to Predict Winner for Emmys

Have you ever watched a show and known right away that it was bound to win the Emmys? Maybe you've been called the Nostradamus of awards, and you have a reputation to uphold to always predict correctly. Or maybe you've been ridiculed your whole life by your friends that your predictions are always wrong, so this time you want to ensure that your prediction has the highest chance of coming true… If you find yourself in any of these situations, fear not as natural language processing (NLP) and sentiment analysis is here to save the day! By leveraging the power of VADER's sentiment analysis package, you have the chance to create a predictive model that can help you impress your friends (and data science professors) while teaching you the foundations of NLP.

Deliverables:

This case study requires you to create a model that will attempt to predict the winner of the next Emmys. To do so, you must scrape reviews from credible websites such as IMDb and Rotten Tomatoes and run those reviews through Python's 'VADER' package, which will quantify those reviews into scores. The reviews you collect must be comprehensive and represent the majority of reviews for a given show. A good project will accurately analyze the sentiment of a review and produce a score that is positive (indicating positive sentiment), or negative (indicating negative sentiment), thereby giving you values needed to create a predictive model of your choice to guess which tv show has the highest chance of winning an Emmy. You will deliver a written report and reflection, as well as a GitHub repository link containing all scripts and resources used for the project. Best of luck and have fun! (Bonus points for correct predictions).