# Online Physician Reviews - Text Mining

# 

# Aneri Shah, Sanghamitra Muhuri, Sujai Karunakaran, Varun Maheswari

# 

# ABSTRACT

## **Background:** Online physician rating websites provide an open-end textual platform for patients to evaluate physicians based on their experiences. The purpose of this project is to identify and evaluate key components of physician reviews which have the strongest effect on the overall rating.

## **Objectives:**

•To examine the relationship among the ratings give to OBGYNs and their reviews

•To determine the sentiment of all the reviews and identifying the impact of each bucket on the doctor’s rating and their average sentiment scores

•To assess if the Ratings or their review sentiments differ on the nature of their age, gender, number of words used in the reviews, experience, etc.

## **Methods:** The Initial dataset was scraped from two physician review websites: Ratemds and Healthgrades. Seven major topics were identified from the reviews of each of the physicians. The reviews were then split into phrases using the NLTK word tokenizer. Using Textblob, sentiment for each of the phrases were identified. Then using a semi-supervised Latent Dirichlet Allocation topic modeling approach, the proportion of each topic in the reviews for each physician was calculated. The dependent variables such as Average Positive Sentiment Score, Average Negative Sentiment Score, Proportion of Positive Score, Proportion of Negative Score, Overall Sentiment was calculated thereby creating the dataset which contains the demographic information of each physician, proportion of positive and negative topics in the reviews and the dependent variables. Then each of the independent and dependent variables was analyzed. The features that greatly affect each of the dependent variables were identified using Multiple Linear Regression.

## **Results:** From the initial analysis of the reviews, the most frequent topics spoken about in the reviews were Medical Expertise, Bedside Manners, Office Staff, Office Environment, Ease of Scheduling, and Waiting time. Similarly, using a semi-supervised Latent Dirichlet Allocation approach the proportion of each of the topics in each of the reviews for each doctor was identified. Among the seven topics, bedside manners and medical expertise were predominantly spoken about in the reviews followed by office staff and waiting time. It was also found that there were more positive reviews than negative reviews. The Exploratory feature analysis revealed that every gender had a significant impact on each of our dependent variables. Namely, Male physicians had better star ratings and positive sentiment when compared to female physicians. Similarly, female physicians had a higher proportion of negative reviews and negative sentiment scores when compared to male physicians. In both the websites, Ratemds and Healthgrades, It was also identified that when the age of the physician increases, the bedside manners decreases irrespective of the age. Although, this trend is not identified with medical expertise. The star rating for each physician decreases when the average number of words per review increases. Followed by the exploratory data analysis, linear regression was used to identify the significant features which directly affect the Star Rating, Average Positive Sentiment, Average Negative Sentiment, Proportion of Positive Reviews, Proportion of Negative Reviews and Overall Sentiment. For Ratemds, it was seen that the star rating was increased by a unit of 0.197 if a physician is a male which aligns with the initial exploratory analysis. Star rating decreased by a unit of 0.004 if the number of words in the review increases. Among the topics, the positive bedside manners greatly contributed to the increase or decrease in star rating at the same time positive waiting time seemed to not affect the star rating. The negative representation of various themes in the review greatly affects most of the target variables than the positive representation of themes in the review. Although similar trends were identified for Healthgrades, for star rating, the positive themes contributed more when compared to the negative themes. The data for Healthgrades and Ratemds was then combined to perform multiple linear regression. The trends in the result were similar to the result we obtained for two separate websites.

## **Conclusions:** Even though the gender and experience of the physicians are significant dependent variables, a physician looking to improve their service can do nothing about these two features. Having said that, the scope of improvement is evident in the case of the topics that were frequently identified in the reviews. A physician should look to have top-notch bedside manners since this area affects each of the dependent variables greatly. This is because, from our analysis, we can see that even when the medical expertise is positively spoken about, the negative bedside manner can decrease the star rating significantly and increase the average negative sentiment. From our analysis, we can see that the waiting period, ease of scheduling, and office staff are comparatively less significant in driving the star rating whereas bedside manners, medical expertise, and cost are the most important variables.