

Twitter-Driven Analysis of *King Richard*: Reception and Engagement

Anthony Porporino (260863300), Yuri Sarasty (260295835), Chris Spennato (260659466)

McGill University

Abstract

Topic and sentiment analysis are useful tools to evaluate discussion on social media. Additionally, TF-IDF scores can give insight into the most significant language in online discussion. In this paper, we use these tools to highlight the salient topics discussed regarding the movie *King Richard*, the relative engagement with the topics of discussion, and the overall perception of the film. Using a sample of one thousand tweets collected in the three-day period following the theatrical release of *King Richard*, our findings show general positive regard for the film and its lead actor, Will Smith, while also noting concern over production decisions concerning representations of race and gender.

Introduction

King Richard details the life of Richard Williams, the domineering father and coach of famous tennis players, Venus and Serena Williams. The movie was released on November 19, 2021.

Our project used topic and sentiment analysis to understand Twitter discourse concerning the theatrical release of *King Richard*. The goal was to highlight the salient topics of discussion, to reveal the relative engagement with those topics, and to gauge the overall perception of the movie.

To begin, we collected a sample of one thousand tweets. We gathered the collection over a period of three days following the theatrical release of the movie. The tweets selected all included the keyword "king richard", used the hashtag #kingrichard, or addressed @KingRichardFilm.

Using our collection of raw data, we then conducted an open coding, and used the results to design a typology of eight topics. We manually annotated each entry in the data-set with both an assigned topic and a label indicating sentiment. Sentiment was labelled as positive, neutral, or negative. We also ran a python script to extract the top ten words with the highest TF-IDF scores in each of our created topics.

Our designed topics indicate the salient topics of discussion on Twitter, while the distribution of these topics help specify the relative engagement. Furthermore, our collection of words with the highest TF-IDF scores provide insight into the vocabulary of discussion for each topic. Our findings provide understanding into the nature and sentiment of

Twitter discourse immediately following the release of the movie.

The most noteworthy topics we observed include: praise for the film; praise for Will Smith; and concern that the film should have focused on the Williams sisters instead of their father. As well, the overall perception of the movie was very positive with 56.7% of the tweets containing a positive sentiment, while only 4.8% contained a negative sentiment.

Data

Our data-set consisted of one thousand tweets, collected using the Twitter API over a period of three days from November 19, 2021 to November 21, 2021. We collected tweets at five instances over the three days: at 12:00 EST and 16:00 EST on November 19 and 20, and 12:00 EST on November 21. In total, we collected 200 tweets on the twenty-first, and 200 tweets at each collection time on the twentieth. On the nineteenth, 298 tweets were collected at 16:00 EST and 102 tweets were collected at 12:00 EST. At each instance, we collected the most recent Tweets that matched the keyword "king richard", included the hashtag #kingrichard, or mentioned @KingRichardFilm, the film's official Twitter account. Only English language tweets were considered. We did not include retweets and replies in the collection to avoid double counting tweets.

Methods

From our data-set, we randomly selected 200 tweets to conduct an open coding. A detailed description of this typology appears in the Results section. Our topic categories included specific descriptions and examples so that each team member could annotate the data-set without relying on individual subjective interpretations.

Our team used Microsoft Excel to annotate the data, assigning each tweet one topic category. At this time, we also labeled each tweet with a sentiment: either positive, negative, or neutral.

We further analyzed the data by computing the term frequency inverse document frequency (TF-IDF) scores for all words in each category. The TF-IDF score for a word-category pair is calculated using the formula shown in Figure 1.

To ensure our TF-IDF score only considered meaningful words, we then pre-processed the annotated data-set. Our

$$tfidf(word, category, dataset) = tf(word, category) * idf(word, dataset)$$

$$tf(word, category) = \# \text{ of times word appears in Tweets with that category}$$

$$idf(word, dataset) = \log (\# \text{ of categories in dataset} / \# \text{ of categories that have the word})$$

Figure 1: TF-IDF formula

pre-processing method consisted of first removing all non Unicode characters (this was mostly emojis), as well as setting every token to lowercase. We then removed all punctuation and Twitter tokens such as: -, @, and #, replacing them with a space character. This measure ensured that we only analyzed actual words from the tweets and that words are easily separable by a white-space character.

We wrote a python script to calculate the TF-IDF score for each word in each topic category. The script extracted the top ten words with the highest TF-IDF score for each category.

We then used the output of this script to remove stop words so that our results only included words which were meaningful to our analysis. The stop words removed include:

“and,” “or,” “a,” “was,” “is,” “really,” “she’s,” “e,” “her,” “his,” “shes,” “go.”

We chose these words because they appeared in the original output, but did not convey meaningful information about the topics nor the general Twitter discourse.

After removing stop words, we were then able to run our final TF-IDF score script. The final output of our script can be seen in the Results section, Table 1.

Results

Our typology consists of eight topics designed to gain insight into the ideas discussed in the data-set, as well as to avoid ambiguity and minimize subjectivity during manual annotation. The eight topic categories and their descriptions include:

opinion:

This category includes all tweets that express positive or negative opinions about the movie. This includes praise, criticism, recommendations, as well as any general review of the movie as a whole.

planning or watching the movie:

This category includes all tweets that express the intention of watching the movie, or indicate that the author of the tweet is currently watching the movie.

family williams:

This category includes all tweets that mention the Williams sisters or Richard Williams in real life, or any comment on their portrayal in the movie.

will smith:

This category includes all tweets that mention the actor Will

Smith as the focus of the tweet. Comments on Will Smith’s acting in his portrayal of Richard Williams would be categorized here rather than in “family williams” category.

family:

This category includes all tweets that discuss the qualities and characteristics of the Williams family. Tweets in this topic category consist of personal comparison and reflection - relating the Williams family to the author’s own. This topic also includes any general statements about raising children or what constitutes a good family.

social concerns:

This category includes all tweets with any critical discussion of social concerns in reference to the movie. This includes assertions of racism or sexism in the movie or its production, and counterclaims that reject that premise.

box office:

This category includes all tweets that discuss the movie’s box office performance.

other:

This category includes all tweets that are related to the film, but have an unclear intention and do not fit well into any other described topic.

Figure 2 shows the distribution of tweets belonging to each category.

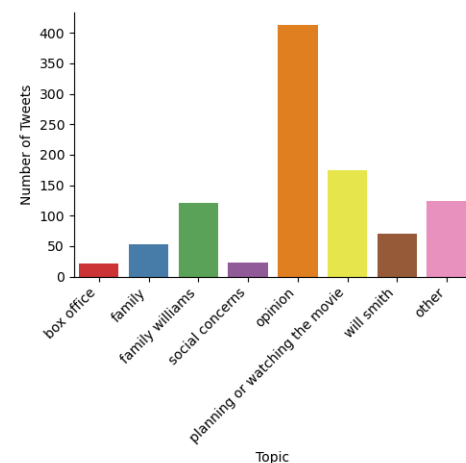


Figure 2: Tweet Counts by Topic

Figures 3, 4, and 5 show the distribution of tweets by topic for each sentiment

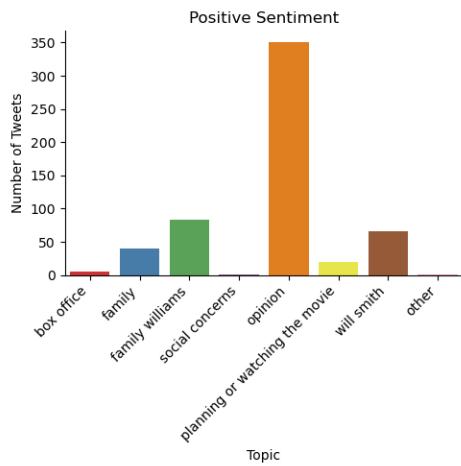


Figure 3: Positive Tweets by Topic

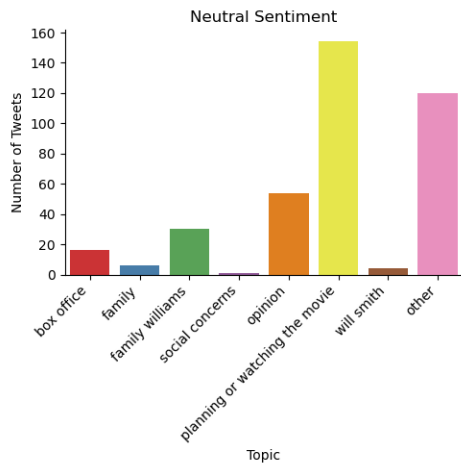


Figure 4: Neutral Tweets by Topic

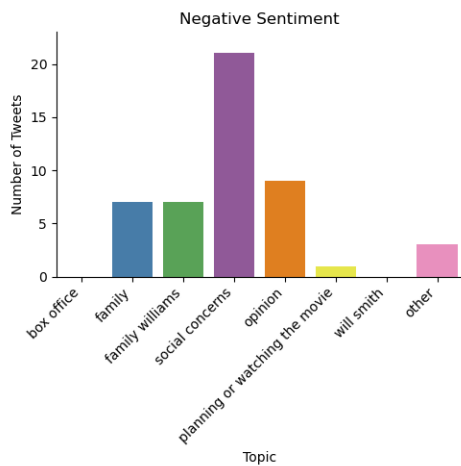


Figure 5: Negative Tweets by Topic

Figure 6 provides a breakdown of tweets categorized as "opinion", by their labeled sentiment.

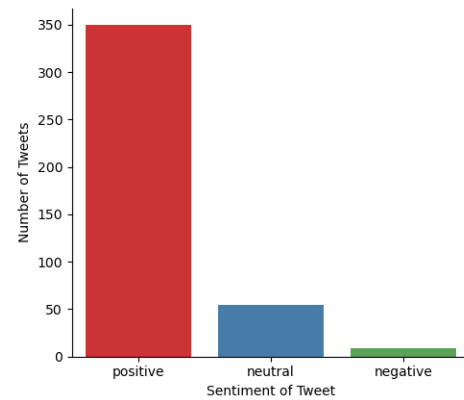


Figure 6: "opinion" Tweets by Sentiment

Table 1 shows the top words in each category by TF-IDF score.

Discussion

Our results foremost show that the film *King Richard* had very positive reception immediately following its theatrical release. Out of one thousand tweets, 56.7% held a positive sentiment, and only 4.8% negative, as seen in Table 2. Furthermore, all TF-IDF words in the "opinion" category had a positive connotation, such as "10/10," "enjoyed," and "great." As well, Figure 2 shows that "opinion" was the category with the most tweets, while Figure 6 demonstrates that "opinion" tweets were overwhelmingly positive in sentiment. Therefore the discussion was generally of positive opinion. Additionally, as seen in Table 3 and Figure 7, the percentage of tweets that had a positive sentiment remained consistently high, indicating opinions did not change drastically as more people saw the movie over its opening weekend. Taken all together, this is a major indication that the overall response to the film was positive.

Moreover, no tweets in the "will smith" category had a negative sentiment, as seen in Figure 5. This indicates that most twitter users had a positive reaction to his performance in the lead role. As seen in Table 1, the top TF-IDF words for the "will smith" category included the word "oscar." This language likely signifies that many viewers found his performance outstanding and deserving of an Academy Award. Interestingly, the last time Will Smith was nominated for an Academy Award was 2007, so this could indicate a return to form for Smith.

The final significant trend in our findings was an apparent concern over the representations of race and gender in the film. As seen in Figure 5, "social concern" was the category with the most tweets labeled as negative. Additionally, the top TF-IDF words from this category include: "white," "feminism," and "women," see Table 1. The vocabulary and sentiment we identified reveal a group dissatisfied with the film's choices when depicting the story of two influential

Topic	Top Ten Words by TF-IDF Score
opinion	good, great, highly, 10/10, movie, times, enjoyed, outstanding, such, oscar
planning or watching the movie	tonight, watch, weekend, wanna, watching, gotta, later, tick, excited, baby
family williams	wanted, venus, serena, beads, break, andserena, sisters, younger, real, williams
will smith	smith, will, smith's, oscar, performance, makes, review, emmanuel, appreciated, hell
family	fatherhood, kids, sacrifice, parents, fathers, aunjenue, raison, papa, themes, highlight
social concerns	white, women, feminism, feminist, claiming, thoughts, appreciate, feminists, wanted, successful
box office	million, \$44, office, boxoffice, \$5, box, 5m, 7m, \$2, whiffs
other	motion, original, promote, picture, harris, kamala, matter, lavar, alive, ball

Table 1: Top Words by TF-IDF Score

Black women. This finding could be interesting to explore further on its own, as well as reactions to this perceived disappointment.

negative	neutral	positive
4.8%	38.5%	56.7%

Table 2: Total Sentiment (%)

date	negative	neutral	positive
11/19/2021	2.50%	45.50%	52%
11/20/2021	5%	33.25%	61.75%
11/21/2021	9%	35%	56%

Table 3: Sentiment by Day (%)

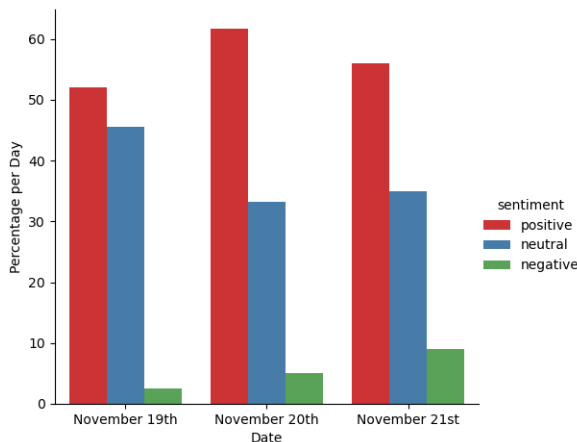


Figure 7: Sentiment by Day

Although we were able to determine some insights, realistically our data-set was quite limited; our sample of one thousand tweets over a three-day period was useful to identify overall trends, but is still quite small. Further work using

a larger sample of data would be useful to uncover more information on the public's engagement with the film.

To conclude, the top three salient topics of discussion regarding the film *King Richard* on Twitter immediately following its theatrical release are: praise for the film overall; praise for Will Smith; and concern over production decisions regarding the representation of race and gender. Overall we can confidently confirm that the movie was very well-received and enjoyed by viewers.

Contributors

To manage group collaboration, we used the online tool Asana. We divided the project into thirteen tasks and assigned duties evenly between group members. We also set up bi-weekly online meetings to collaborate and stay on track. Even though only one person completed each task, all three team members reviewed every component of the project. Therefore this report is the result of both individual work and collective collaboration. Each team member contributed valuably and equally.

Chris handled tweet collection, with input from the team on details concerning keyword selection and time selection. All team members participated in open coding and annotation of the tweets. Yuri and Anthony calculated the TF-IDF scores, while all members helped compile the final report.