MI1 Sea Monsters Checkpoint
Group: The Sea Monsters
Group Leader: Audrey Himes Group Members: Alexa Owen, Sharaf Tariq
DS 4002
February 9, 2023

Narrative: Most people watch movies and think to themselves, “That was a great movie!”, or “That movie was not worth my time.” However, few actually write reviews on the internet about the film after watching it. Despite the fact that fewer people write reviews than watch the movie, people use the reviews written by critics and peers to understand if they want to watch a certain film. Numerical ratings can only show whether people liked a movie or not, but they do not provide any context to *why* people may or may not like it. People write reviews to give their opinions on movies, which others read to analyze if they want to watch them [1]. Even the Academy values movie reviews, as they are partnering with Letterboxd, a movie reviewing app, to create content for the Oscars [2]. Furthermore, numerical ratings remove the nuances of what people like about movies. There are many movies that are ‘bad’, but people still love them. This cannot be properly understood when solely looking at numbers. Written movie reviews are important to get a full picture of a movie without spoiling too much of it.

Hypothesis: There is a connection between movie review sentiment and the ultimate rating that the film receives.

Research Question: Can the sentiment of the language used in a movie review be related to a numerical rating from 1 to 10?

Modeling Approach: We will be using natural language processing and text mining to conduct a sentiment analysis. Our main approaches will be to use the National Resource Council emotion lexicon (nrc) algorithm, afinn analysis, and bing analysis to understand the sentiment within our text data. Each of these methods provides a slightly different manner for dissecting and understanding the sentiment of our text. The nrc analysis will help us to understand the emotions behind the language in our data, while the afinn and bing analysis methods will provide us with a more basic positive/negative distinction for each word in the text [3]. We want to explore both of these routes to understand how both emotional and more objective language factor into creating an overall numerical rating within a film review. It is possible that each of these methods leads to slightly different conclusions about our data, so it is important to explore them all in order to develop a holistic understanding of the relationships between language sentiment and numerical ratings.

References:

[1] S. Renshaw, “Reflections on 30 years of film criticism”, *Salt Lake City Weekly*, para. 5, Feb. 8, 2023. [Online], Available: https://www.cityweekly.net/utah/reflections-on-30-years-of-film-criticism/Content?oid=19549447. [Accessed Feb. 9, 2023]

[2] J. Tangcay, J. MacCarey, and C. Flam, “The Academy Partners With Letterboxd to Create Content for the 95th Oscars – Film News in Brief”, *Variety*, para. 2, Jan. 23, 2023. [Online], Available: https://variety.com/2023/artisans/awards/film-news-in-brief-jan-23-1235499284/. [Accessed Feb. 9, 2023].

[3] J. Silge and D. Robinson, “Sentiment analysis with tidy data”, *Text Mining with R,* Newton, MA: O'Reilly Media, 2017. [E-book] Available: https://www.tidytextmining.com/sentiment.html.