**Oppenheimer Movie Sentiment Analysis**

Kankana Ghosh

**Introduction:**

In the modern digital age, the entertainment industry is closely intertwined with the sentiments and opinions expressed by audiences. Movies, as a form of art and entertainment, evoke a wide range of emotions and reactions among viewers. Understanding these reactions is crucial for filmmakers, producers, and distributors to tailor their strategies, enhance audience engagement, and optimize marketing efforts. Sentiment analysis, a prominent application of natural language processing (NLP), offers a powerful tool for extracting insights from textual data to gauge public sentiments. The objective of this project is to conduct sentiment analysis on movie reviews to decipher the overall sentiment of the audience towards different movies. By applying advanced data science techniques to a corpus of movie reviews, we aim to unveil patterns and trends in sentiment expression, providing valuable insights into audience perceptions.

**Methodologies:**

The dataset contains 84048 observations on 2 variables. It doesnot contain any missing values but have many duplicate rows. After removing the duplicated rows the new dataset contains 2036 observations on 2 variables. Firstly fitted the TextBlob Model. One of the primary features of TextBlob is sentiment analysis. Sentiment analysis involves determining whether a piece of text expresses positive, negative, or neutral sentiment. TextBlob provides a built-in sentiment analysis function that assigns a polarity score to a given text, indicating its sentiment. The polarity score ranges from -1 (negative sentiment) to 1 (positive sentiment), with 0 being neutral. After getting the TextBolb scores and sensitivity I added sentiment label indicating whether the review given by the viewer is positive, negative or neutral. Then I constructed a plot of Textblob score distribution across rating categories. By inspection it’s clear that maximum viewers have given positive review about Oppenheimer. Now, implementing the VADER (Valence Aware Dictionary and Sentiment Reasoner) which is particularly effective at capturing sentiment nuances, including positive and negative sentiments, neutral sentiments, and even intensity modifiers. Using the VADER model I got the compound, positive, neutral and negative scores and also added sentiment labels. The barplot for the number of sentiment labels using VADER model also gives the same result as TextBlob model, i.e., maximum viewers have given positive reviews about the movie.

**Conclusion:**

In conclusion, the sentiment analysis of movie reviews project showcases the convergence of data science and the entertainment industry. By decoding the emotional tapestry woven within movie reviews, this project exemplifies the potential of data-driven insights in shaping the cinematic experience. Through sentiment analysis, the world of movies becomes not only a canvas for artistic expression but also a domain where technology facilitates a deeper connection between creators and their audience. The TextBlob model is a great choice for quick and simple tasks but it may not provide the same level of accuracy as other models. On the other hand, VADER model is a valuable tool for sentiment analysis in social media and informal text contexts. Its lexicon-based approach and focus on sentiment intensity make it particularly well-suited for capturing sentiments in short and expressive text data.