Github link for Final Deliverables: https://github.com/UjwalGupta12/INST737-Project-

INTRODUCTION

For our INST737 team project, we chose to work on a sentiment analysis comparing Star Wars movies I to VI and movies VII to IX. Our motivation for choosing Star Wars as our topic was because we are both big Star Wars fans who watched many Star Wars media. We also decided to choose the movies, and not the shows, because they are well known, and we are curious in wanting to understand what the other fans and general public thought about the sequel trilogy movies. Our purpose for conducting this sentiment analysis is that we hope to provide our potential clients, Disney, with results on comparing the sentiment for movies I to VI and movies VII to IX. And that by doing so, we can inform Disney on what made the movies successful or not and to hopefully be applied to future movies.

BACKGROUND

Star Wars I to VI were by George Lucas under Lucasfilm and Star Wars VII to IX were created by Disney. Star Wars IV was released in 1977, Star Wars V in 1980, and Star Wars VI in 1983. These three movies are known as the prequel trilogy. Star Wars I was released in 1999, Star Wars II in 2002, and Star Wars III in 2005. These three movies are known as the original trilogy. George Lucas sold Lucasfilm to Disney in 2012. The sequel trilogy would be made, with Star Wars VII being released in 2015, Star Wars VIII in 2017, and Star Wars IX in 2019.

Summary and Contributions

A summary of the project can be outlined as follows: a) web scraping Star Wars I to IX from IMDb; b) performing topic modeling; c) performing sentiment analysis on each movie; d) divided sentiment analysis by positive and negative sentiment; and e) sampled reviews and found reasons for the positive and negative reviews. (Figure 1)

On this project, both Matthew and Ujwal worked on web scraping the IMDb reviews for Star Wars I to IX, topic modeling, and sentiment analysis. Matthew analyzed the sampled negative reviews and Ujwal analyzed the sampled positive reviews.

LITERATURE REVIEW

When talking about Star Wars movies, one of the best ways to learn about how well each movie did is through rankings. From IGN, a multimedia site that does reviews, the IGN Staff compiled a ranking of the Star Wars movies from worst to best. The list does not include every movie, but the order is: *Star Wars: The Rise of Skywalker*, *Star Wars: Attack of the Clones*, *Star Wars: The Phantom Menace*, *Solo: A Star Wars Story*, *Star Wars: Revenge of the Sith*, *Star Wars: The Force Awakens*, *Rogue One: A Star Wars Story*, *Star Wars: The Last Jedi*, *Star Wars: Return of the Jedi*, *Star Wars: A New Hope*, and the best being *Star Wars: The Empire Strikes Back* (IGN, p. 3-10). This aspect of focusing on the movies from Star Wars and comparing them is what we wanted to explore deeper.

In the study done by Balogh et. al, they analyzed the conversations between characters in the prequel and original trilogy, as well as *The Force Awakens*, and created relationship diagrams with nodes and edges (p. 2-5). They also looked at the sentiment of words exchanged between characters. The aspect on the positive and negative sentiment on the words spoken for movies four to seven is also what we wanted to inspect.

In another study analyzing the script from Star Wars movies four to six by Xavier Garcia, he first analyzed the frequency of dialogue by each character in each movie into word clouds. He also conducted sentiment analysis using lexicon bing to sort words into positive and negative categories, and also used nrc to sort words into categories like: "positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust" (Garcia, Section 8). Xavier also did a sentiment analysis on the frequency of words related to the characters using tf-idf. The aspect of focus in this study is the frequency of words found in each movie using sentiment analysis.

In a sentiment analysis study by Juraj Kapasny on Star Wars movie *The Force Awakens*, he wanted to understand the Twitter user's sentiment on the movie between December 4th to December 29th in 2015. Juraj mapped the location of the tweets on a map, visualized the frequency of tweets in the studied time frame, and also graphed the tweets by sentiment and their location (Kapasny, p. 2-3). The aspect from this study to focus is the sentiment analysis of people's review on the movie.

Overall, our work will be a culmination of the aspects from these literatures. We will be focusing particularly on movies from Star Wars, covering and comparing all the episodes I to VI and VII to IX. We will also be conducting sentiment analysis with topic modeling, which is similar to finding the frequency of the words in the movies, but also grouping the words by topic. Afterwards, we will collect our data not from tweets, but similarly from reviews about what the users thought of the movie from IMDb. As there has already been sentiment analysis mainly on Star Wars IV to VII, we wanted to expand the analysis to all three movies in the prequel and sequel as well. And finally, we wanted to go beyond not just finding positive and negative reviews, but also analyzing the reasons for those sentiments in the user reviews.

Methodology
Data Collection

We collected our data on reviews by web scraping from IMDb using a Python library called Selenium. We chose Selenium instead of BeautifulSoup because Selenium allowed us to click the 'Load More' button. This in turn permitted us to web scrape the subsequent sets of reviews.

Data Cleaning

For data cleaning, we simply removed unnecessary columns of data for our analysis. This included information on the reviews' date, author, rating, title, and URL. We removed these because our analysis focused only on the reviews itself.
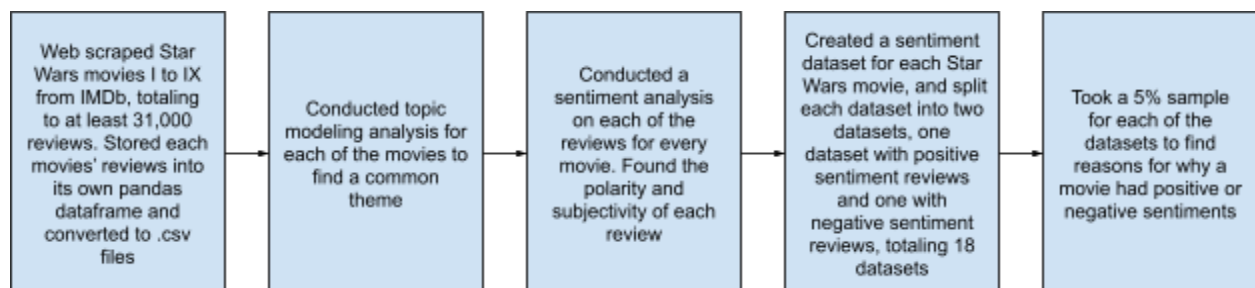
Solution Framework

Figure 1: Solution Framework Pipeline for solving problems on sentiment analysis for Star Wars I to IX.

Experiment Result
Experiment Setup

        The first part of our experiment was topic modeling. To start, we first removed stopwords for each review by using the Python library NLTK. We also removed any new line characters and single quotes with regular expressions. Next, we developed bigram models using the Python library Gensim. Lastly, lemmatized the sentences. We planned to evaluate the results of our topic modeling analysis by examining common patterns in the topics generated as part of the sentiment analysis.

        The second part of the experiment was the sentiment analysis for every review in each of the nine Star Wars movies. For this, we used the Python library TextBlob in order to determine the polarity and subjectivity of each review. After this, we planned to evaluate these results by grouping the reviews based on its polarity score. Since the score goes from -1 to 1, with reviews closer to -1 are more negative, whereas reviews closer to 1 are more positive, we created a for loop that split on polarity and created a dataset for each, totaling 18 total datasets, nine for each positive and negative. We took a 5% sample of each positive and negative dataset and analyzed each review. Splitting the reviews dataset on polarity helped in separating the reviews for easier analysis for reasons of positive or negative reviews.
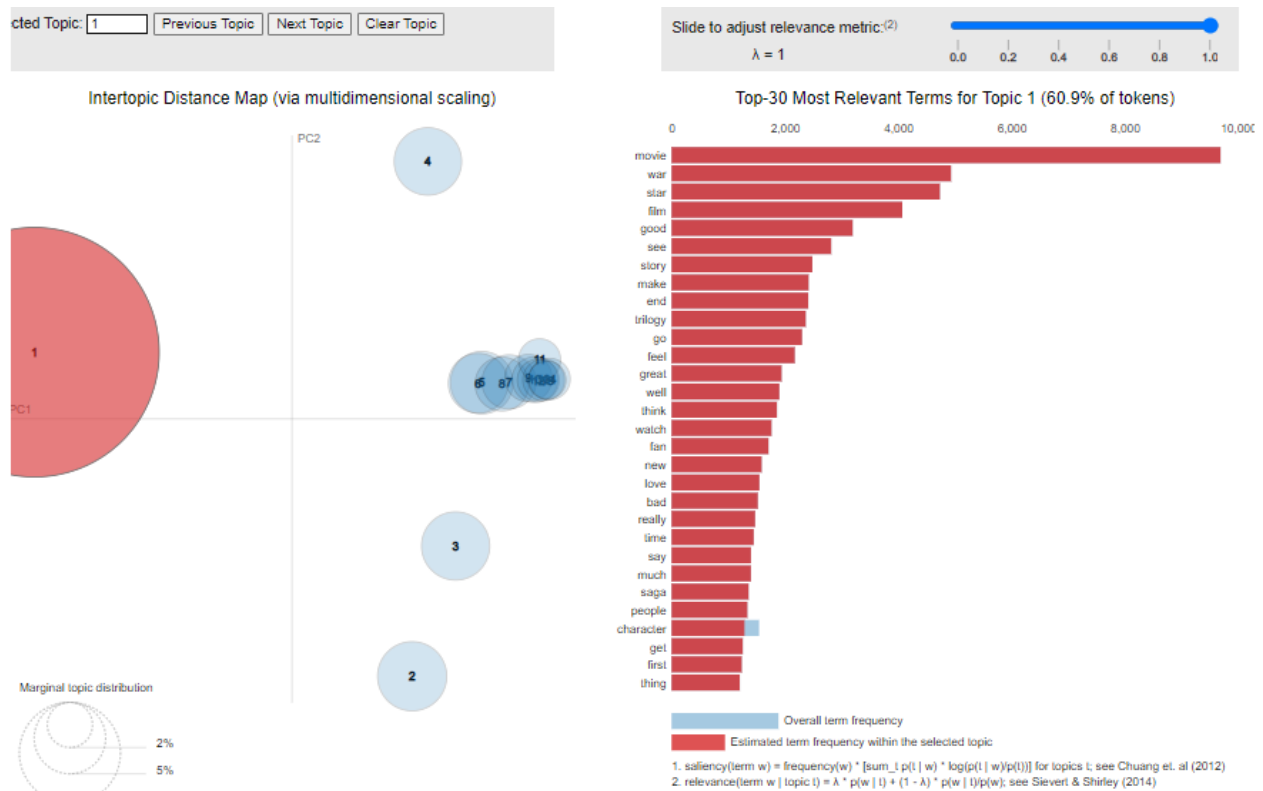
Results

Figure 2: Topic modeling results for the reviews of Star Wars Episode IX: The Rise of Skywalker.

Positive

1		Review	Polarity	Subjectivity
2	1799	This movie is a joke and to be honest this whole trilogy needs to be wiped from the canon because it's a stain on the franchise.	0.4	0.65
3	7104	This movie is just perfect ! Good plot, twists, laugh, action, it got it all. Seen it twice and can't wait to see it again. Si not trust the hatters, they know nothing about the force	0.525	0.45

Negative

1		Review	Polarity	Subjectivity
59	7666	The 'last' (and I emphasise last as there is likely to be another 27) instalment in the Star Wars saga was meant to have it all but instead had all the ingredients of a narrative failure. The obvious, rushed attempt of trying to have it all only highlighted the true essence of the narrative mess.	-0.01770833333	0.4697916667
60	549	Secondarily. Boring. Chaotically. Stupidly. Adam Driver and Daisy Ridley claim to be the worst couple of the year. The script contains so many holes that they can not even plug the Death Star. J. J. Abrams shot a traditional movie gum that loses its taste after ten minutes of viewing.	-0.3714285714	0.6214285714

Figure 3: Positive (Top) and negative (Bottom) excel sheet sample of reviews with polarity and subjectivity for Star Wars IX.

Findings

Episodes	Positive Reviews	Negative Reviews
1	Special effects were liked by 21.7% of reviews	> 50% disliked Jar Jar Binks (Annoying)
2	37% of reviews liked the actions	> 50% dislked weird romance
3	19.5% of reviews liked the actions, some notable scenes that were liked by some reviews were Anakin Vs. Obi-Wan	> 30% disliked bad acting from Samuel L. Jackson and Hayden Christensen
4	40% of reviews liked the overall idea of Star Wars and talking about the large impact the franchise has made	> 30% Movie Overrated and mainly for kids (Silly plot & weak villains)
5	46% liked the plot/story, some had liked the plot twists in the movie	> 30% bad romance, bad acting, inconsistent storyline
6	27.5% liked the plot, notable reasons include Darth Vader's return to being good	60% disliked bear characters, calling it a marketing plot
7	12.5% liked the characters, notable reasons include the return of Han Solo and Leia Organa	> 25% Main actress is a Mary Sue character, lack of character development, so boring characters
8	10% of reviewers liked the visual /special effects	> 20% disliked ruining of Luke Skywalker's character
9	10.3% liked JJ.Abram's directing	>20% Terrible ending and J.J.Abrams ruined sequel trilogy, no longer Star Wars

Figure 4: Shows the most impressive finding for positive and negative review for each of the nine movies.

Limitation

      Our main limitation for this project was time, and consequently knowledge. Due to us only having one semester to work on the project, we were limited with time to learn the Python libraries and code needed. As a result, we were challenged on using unfamiliar Python libraries like Selenium. In addition, due to time, we did not have the chance to fully learn and implement Python libraries spaCy and

NLTK for part of speech tagging to help us in finding reasons. Had we been able to learn this, we would have been able to analyze the reasons for every review in each movie instead of sampling the reviews.

Conclusion and Future Work
Conclusion

We had successfully web scraped the reviews from Star Wars movies I to IX using Python library Selenium, gathering at least 31,000 reviews. Conducting topic modeling on each movie helped us understand the common topics of the reviews. We used Textblob for sentiment analysis, sampled the reviews, and found reasons for positive and negative reviews. Overall, in Star Wars I to VI, the positive reviews summarize that the prequel trilogy (episodes I to III) action scenes and special effects were liked by many. What was most liked about the original trilogy was the plot or story. What was summarized negatively for Star Wars I to VI was due to the poorly written dialogue, excessively bad CGI, silly plot, weak enemies, bad romance, and bad acting from some actors. For Star Wars VII to IX what was positively reviewed summarized to many praising J. J. Abrams for directing episode VI and IX. Many had also liked the return of old characters such as Luke Skywalker, Leia, and Han Solo. The negative reviews for Star Wars VII to IX stemmed from having a repeated storyline, a Mary Sue character, plot holes and unexplained actions, ruining fan favorite old characters, and blaming J. J. Abrams for ruining Star Wars. For our recommendations to Disney when making more Star Wars movies, we would like to suggest the following: a) having a well written, coherent story; b) having a few, interesting characters with ample character development rather than too many uninteresting and unnecessary characters; c) to not reuse similar storyline from previous movies; d) fewer and better jokes; and e) to not upset the fan base by finding the right balance to satisfy old fans while intriguing new audience.

Future Work

When it comes to future work, we want to perform topic modeling and sentiment analysis with more Star Wars related content, such as Star Wars television shows. The reason is because we believe comparing the sentiment between Star Wars television shows and Star Wars movies would be interesting. Another future task would include web scraping from other review sources like Rotten Tomatoes and Metacritic. Comparing reviews between the review sources would also be interesting to analyze for how similar and different they might be. One final future task to implement is incorporating part-of-speech tagging using spaCy and NLTK in order to find reasons for the reviews as well.

References
Balogh, K., Szucs, K., Verecze, V., & Varju, Z. (n.d.) *Visualizing Star Wars Movie Scripts*. Precognox.
        https://tas.precognox.com/labs/star-wars-visualization/#
García, X. (2020, August 26). *Analyzing Star Wars Movie Scripts*. Kaggle.
        https://www.kaggle.com/code/xvivancos/analyzing-star-wars-movie-scripts
IGN Staff. (2022, May 28). *Ranking the Star Wars Movies From Worst to Best*. IGN.
        https://www.ign.com/articles/best-star-wars-movies
Kapasny, J. (2016). *The Data Awakens: Star Wars Sentiment Analysis*. KD Nuggets.
        https://www.kdnuggets.com/2016/01/data-awakens-star-wars-sentiment-analysis.html