**MKTG6010 Assignment 3**

**Final exam**

SID 510332280

Word count: 1448 words

30/05/2023

Table of Contents

**Introduction3**

**Question 1: topic modelling3**

**Question 2: sentiment analysis5**

**Question 3: regression analysis6**

**Summary8**

**Reference list9**

**Appendix10**

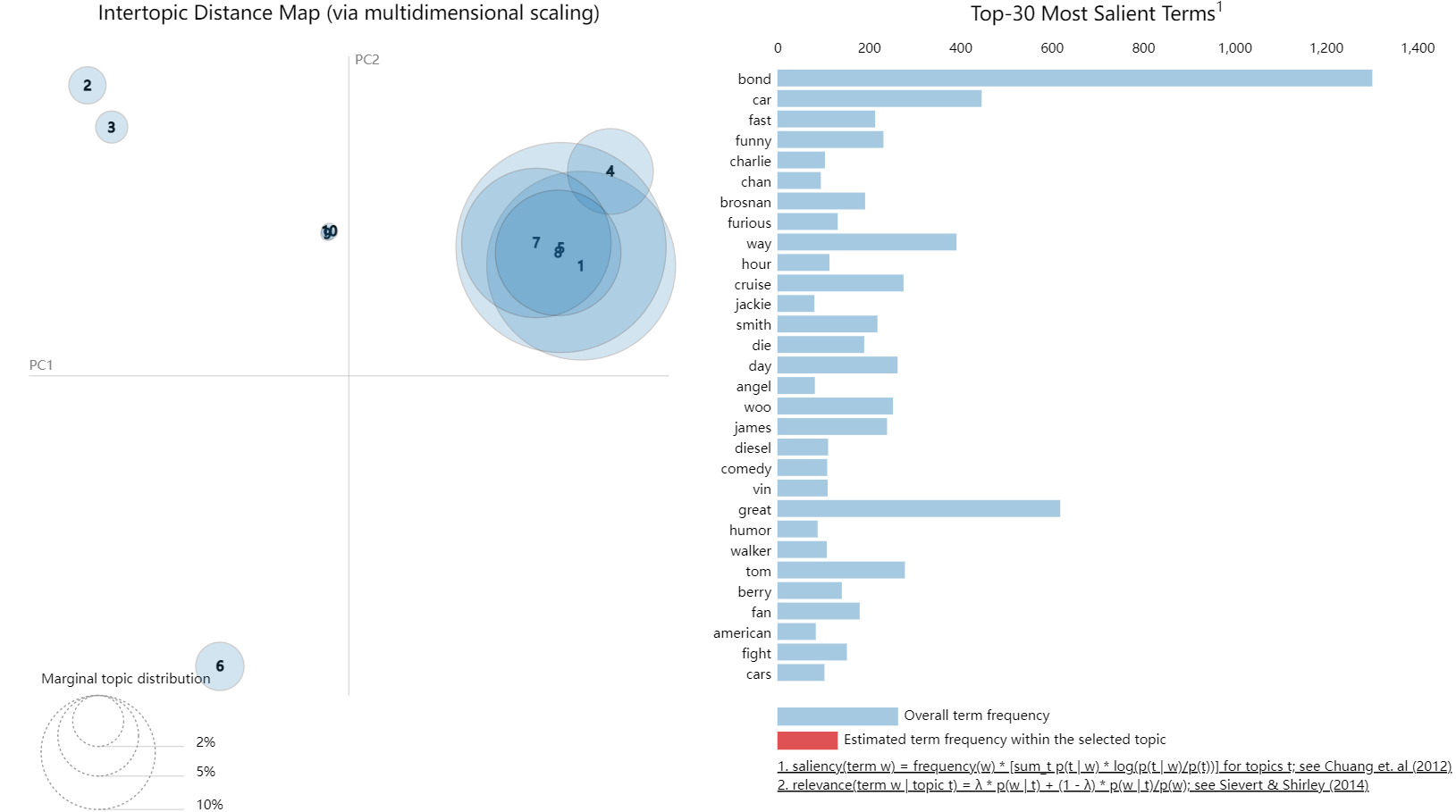
1. **Introduction**

The Internet Movie Database (IMDb) is “an online database containing production details, cast lists, technical details, merchandising information, box-office data, plot summaries, user ratings, and film-related trivia about a wide range of films (Kumar et al., 2019). Nowadays, social media has diversified to understand the experiences of movie fans, including their reviews and responses. In this report from the perspective of a film producer, an analysis of reviews and other factors is conducted using topic modeling, sentiment analysis, and regression analysis. The goal is to examine the impact of various factors on the movie box office and provide recommendations for the film industry and marketers to optimize marketing strategies, increase box office revenue, and enhance fan satisfaction.

The dataset is sourced from IMDB and includes movies of all types. However, the analysis focuses specifically on action and comedy movies. This focus group approach aims to delve into the different varieties of fan engagement and identify potential box office revenue patterns associated with each segment. Before conducting the analysis, it was necessary to address missing values in the dataset. Five rows with missing values were removed to ensure data quality improvement.

1. **Topic modelling**

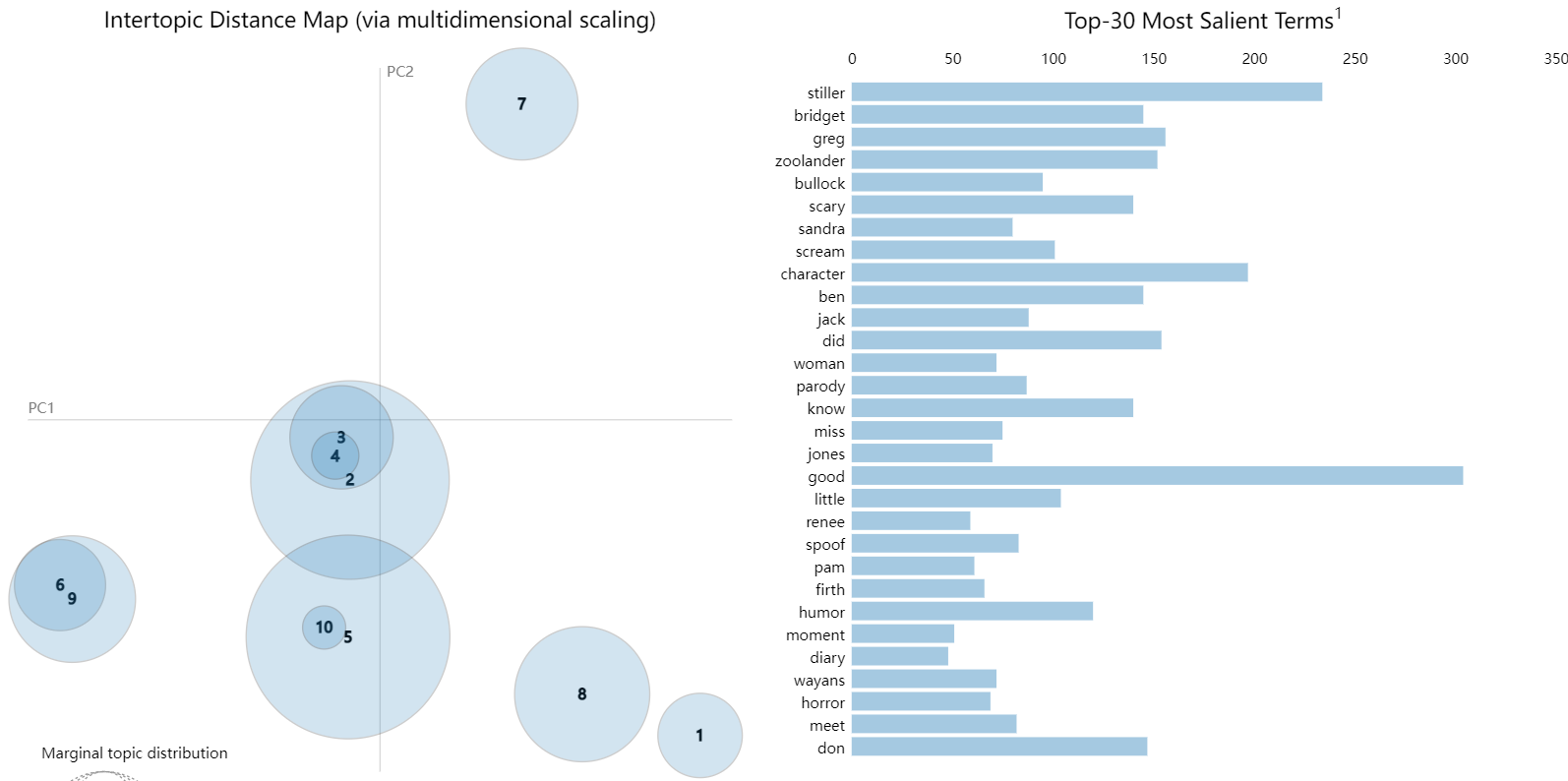
According to appendix 1, keywords associated with action movies mainly include actors, characters, and scenes. Figure 1 shows the overall topic modelling results of action movies. Moreover, topic 5 stands out 33.9% as the key section. Appendix 2 highlights the three key names: Ocean's Eleven, George Clooney, and James Bond. As Emily Garbutt (2003) reported, George Clooney said that there were a couple of other A-listers who could have been on the billing but turned it down including James Bond actor Daniel Wroughton Craig from the 007 series Therefore, this topic garnered significant attention and became the hottest topic of discussion among movie enthusiasts. In appendix 3, topic 1 captures the second-highest proportion at 27.3%. It showcases the immense adoration of fans for Mission: Impossible II, featuring Tom Cruise. The film's success can be attributed to its gripping storyline, enhanced better by adaptations from the original story. Additionally, Tom Cruise's portrayal of Ethan has captivated a broad fan base. As a result, Mission: Impossible II achieved the highest box office revenue in the entire film industry in 2000 (Emily 2003).



*Figure 1. Topic modelling results of action movies*

Figure 2 illustrates the keywords associated with comedy movies, highlighting a distinct focus on the emotional and intellectual impact they evoke, unlike action films. Moreover, terms like "scream," "scary," and "horror" exemplify the range of emotions expressed in comedy movies.

In Appendix 4, topic 5 comprises 26.5% of the importance, and mainly around the film Zoolander. Directed by Ben Stiller, this 2001 comedy blockbuster combines social and cultural references, playing a defining role in shaping modern comedy. Zoolander's inherent stupidity and cultural pertinence are more suitable to the environment of contemporary cinema, aside, perhaps, from the cameo appearance of Donald Trump and for initiating discussions about women's societal roles（Calum 2021）. Many female movie enthusiasts passionately expressed their admiration for the film on IMDB, generating considerable buzz. Appendix 5 explores topic 2 accounting for 25%, which delves into the portrayal of various roles in comedy films. It sheds light on the prevalence of female characters being consistently depicted as "momma" in comedies, perpetuating outdated scenes and stories (Calum 2021). This topic has captured the attention of movie fans, prompting them to contemplate the entrenched nature of these roles.

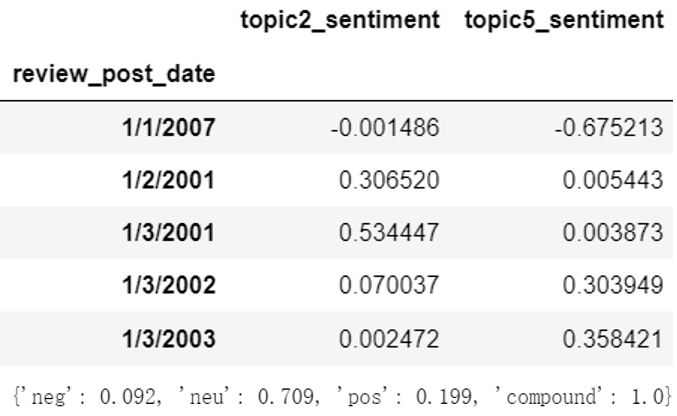
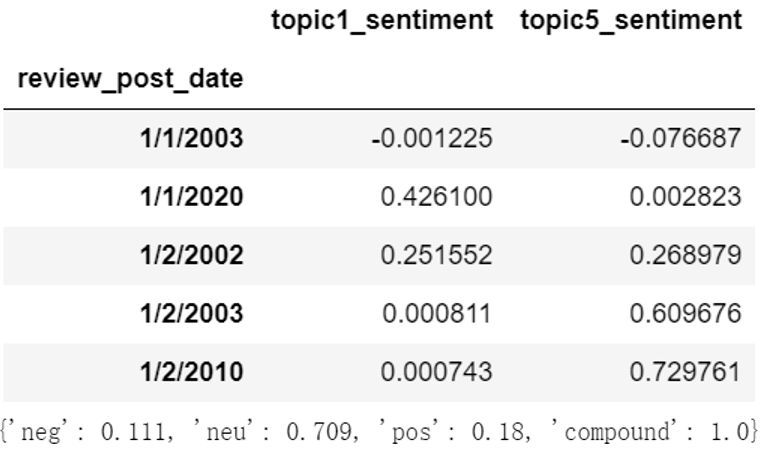


*Figure 2. Topic modelling results of comedy movies*

Based on the analysis, attracting audiences for action movies relies on the presence of popular actors and well-adapted scripts. To capitalize on their popularity, encouraging movie stars to engage social media platforms, and share behind-the-scenes footage, trailers, and exclusive content fosters audience participation, fueling discussions and generating excitement for the film. In the case of comedy movies, audience perception emphasizes the significance of emotions and intellectual stimulation. In marketing efforts, emphasize the emotional expressions and humor depicted in the movie. Utilize impactful trailers and catchy slogans to convey the film's underlying thoughts and spirit. For instance, by addressing gender issues in a movie like Zoolander, connecting these social topics with the film's promotion can attract attention and spark discussions among the audience on various social media platforms.

1. **Sentiment analysis**

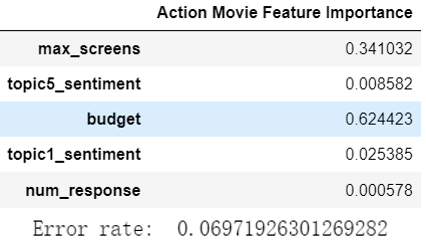
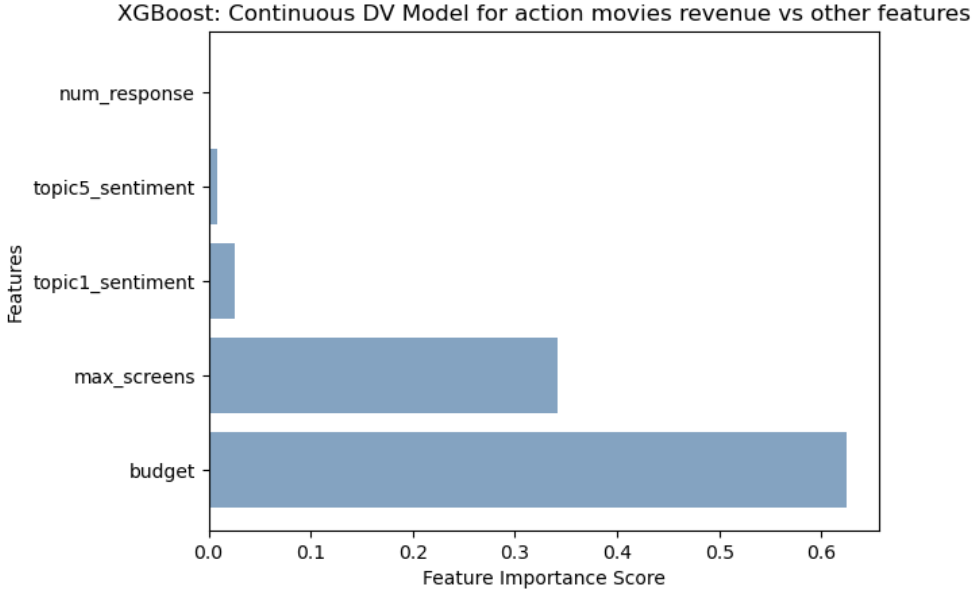
Sentiment analysis is a classification process to understand the opinions, interactions, and emotions of movies (Ramadhan et al., 2022). As shown in Figure 3, action movies have an overall positive sentiment score, with 70.9% neutral, 18% positive, and 11% negative sentiment. Comparing the sentiment scores of topic 1 and topic 5 reveals that topic 5 has a significantly higher sentiment score, indicating stronger acceptance of Mission: Impossible adaptations but remaining uncertain about George Clooney. Figure 4 shows those comedy movies also have a positive overall sentiment score. Interestingly, although both action and comedy movies share a 70.9% neutral sentiment score, comedies have a higher positive sentiment score of nearly 20% and a lower negative sentiment score of 9.2%. This suggests that comedy movies tend to evoke more happiness, possibly due to their scripts and themes. Within comedy movies, topic 5 has a significantly higher sentiment score compared to topic 2, indicating greater appreciation and agreement with unconventional comedies like Zoolander, while finding comedies with fixed character roles dull. Therefore, for action movies， when marketing action movies, the focus should be on promoting the plot and story rather than spreading negative rumors for attention. For comedy films, the emphasis should be on promoting the underlying spiritual essence beneath the plot rather than superficial character portrayals.



*Figure 3. sentiment scoers of action movies Figure 4. sentiment scoers of comedy movies*

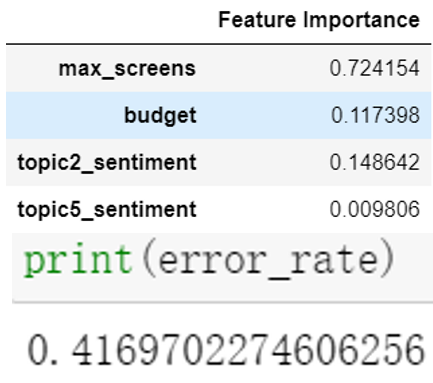
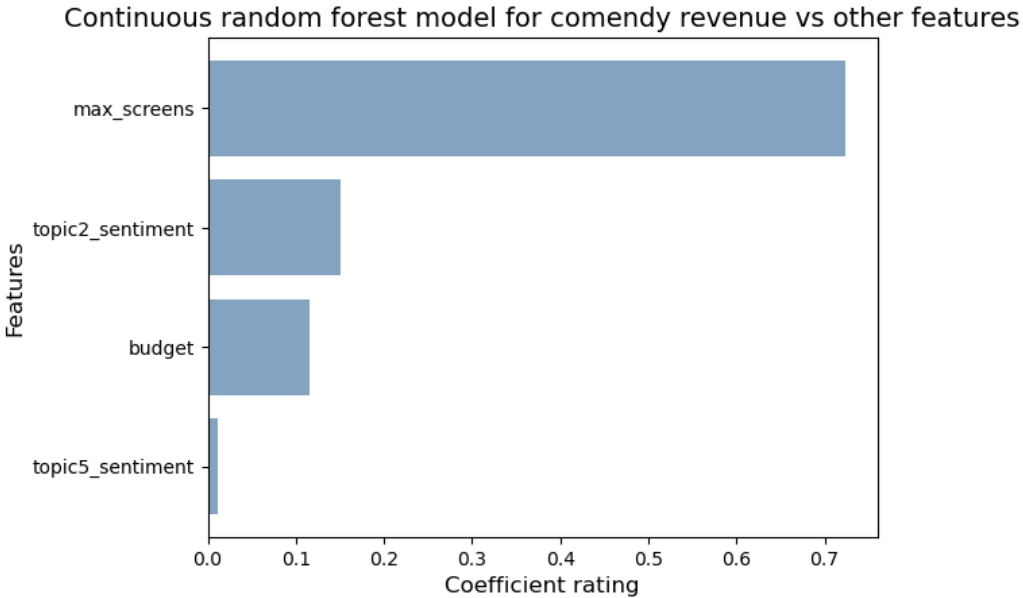
1. **Regression analysis**

For action movies, the linear analysis of action movies in appendix 6 reveals that most factors, including sentiment scores, are statistically significant (p-value < 0.05), except for the number of upvotes for helpfulness on IMDB. However, the model's R-squared of 0.689 indicates that only approximately 68.9% of the variation in box office receipts can be explained by the included features. To validate the accuracy of the OLS model, an XGB boosting model is employed, which is effective in assessing the importance of multiple variables with fewer observations and mitigating the impact of outliers. The results in figure 5 highlight that the budget, accounting for 62.4%, is the most influential factor. Combining it with the OLS model, one unit increase in the budget is associated with a decrease of -0.0002144 units in box office revenue. As movie producers strive to achieve these goals they eventually reach a point where tradeoffs must be made between increasing quality and lowering costs (Farooq et al., 2017). Both sentiment scores contribute a relatively small proportion, possibly due to limited observation values. Specifically, one unit increase in the sentiment score for star-related rumors (topic1) leads to a decrease of 0.0361 in box office receipts. Regarding successful scenes and reputable actors (topic 5), one unit increase in the sentiment score corresponds to a 0.1589 increase in box office revenue. The key strategy for increasing the box office of action movies is to pre-release buzz and star power marketing with attractive scripts. For instance, showcasing trailers on social media prior to the movie's release helps generate audience anticipation and excitement. Conducting positive interviews with movie stars can effectively capture people's attention and convey the film's positive vibes.



*Figure 5. XGBoosting model of action movies*

In Appendix 7, for comedies, except for the number of responses and beneficial votes, other variables, including sentiment scores, show statistical significance (p-value < 0.05). To validate the accuracy of OLS, a random forest model was used as a benchmark. Figure 6 reveals that the most influential factor for comedy movie box office is the number of views, accounting for 72.4%, distinct from action movies. A one-unit increase in maximum views corresponds to a 0.0011 increase in box office revenue. The sentiment score related to the movie's core spirit (topic2) follows, accounting for 14.86%. An increase of one unit in topic2 sentiment score is associated with a -0.2621 decrease in box office revenue. It is hypothesized that comedy's core may be tragic, impacting emotions negatively as more audiences grasp the tragic essence (Calum 2021). Additionally, topic 5 holds minimal importance of 0.009, indicating the limited influence of stereotypical character tasks on the box office. With a coefficient of -0.2351, a one-unit increase in topic5 sentiment score may result in a -0.2351 decrease in box office revenue. The marketing strategy for comedy films primarily revolves around advertising and promotion to increase viewership. It involves creating enticing ticket packages, discounts, and exclusive merchandise to incentivize audiences to watch the movies. Collaborating with theater chains and online ticketing platforms can effectively promote these offers and attract potential customers.



*Figure 6. Continuous random forest model of comedy movies*

1. **Summary**

In general, excellent stories and renowned actors have a greater appeal to audiences in action movies. Pre-release buzz and star power marketing can generate anticipation and boost box office revenue by engaging audiences in social media discussions. Good comedy films transcend superficial humor, captivating viewers with profound themes and societal reflections. To maximize box office earnings, it is recommended to employ concise and captivating advertising slogans.

However, this report has limitations. Firstly, the dataset is from 2000, potentially affecting the accuracy of sentiment scores due to changing values and preferences. Secondly, the small number of observations may result in the underfitting of the OLS model. Future research may focus on collecting recent data for improved analysis.

**Reference list**

Calum, R. (2021，September 14). 20 years of 'Zoolander' Ben Stiller's comedy gamechanger. *Faroutmagazine*.

<https://faroutmagazine.co.uk/zoolander-ben-stiller-20-years-later-gamechanger/>

Emily, G. (2003, April 23). George Clooney reveals the A-list actors who turned down Ocean's Eleven. *Gamesradar*.

<https://www.gamesradar.com/george-clooney-actors-who-turned-down-oceans-eleven/>

Farooq, M. A., Kirchain, R., Novoa, H., & Araujo, A. (2017). Cost of quality: Evaluating cost-quality trade-offs for inspection strategies of manufacturing processes. *International Journal of Production Economics*, 188, 156–166.

<https://doi.org/10.1016/j.ijpe.2017.03.019>

Ramadhan, N. G., & Ramadhan, T. I. (2022). Analysis Sentiment Based on IMDB Aspects from Movie Reviews using SVM. *Sinkron,* 7(1), 39–45.

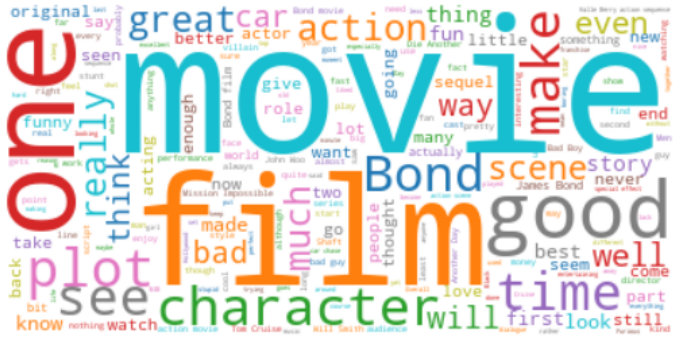
<https://doi.org/10.33395/sinkron.v7i1.11204>

Kumar, H. I. K., Harish, B. S., & Darshan, I. I. K. (2019). Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(5), 109–114.

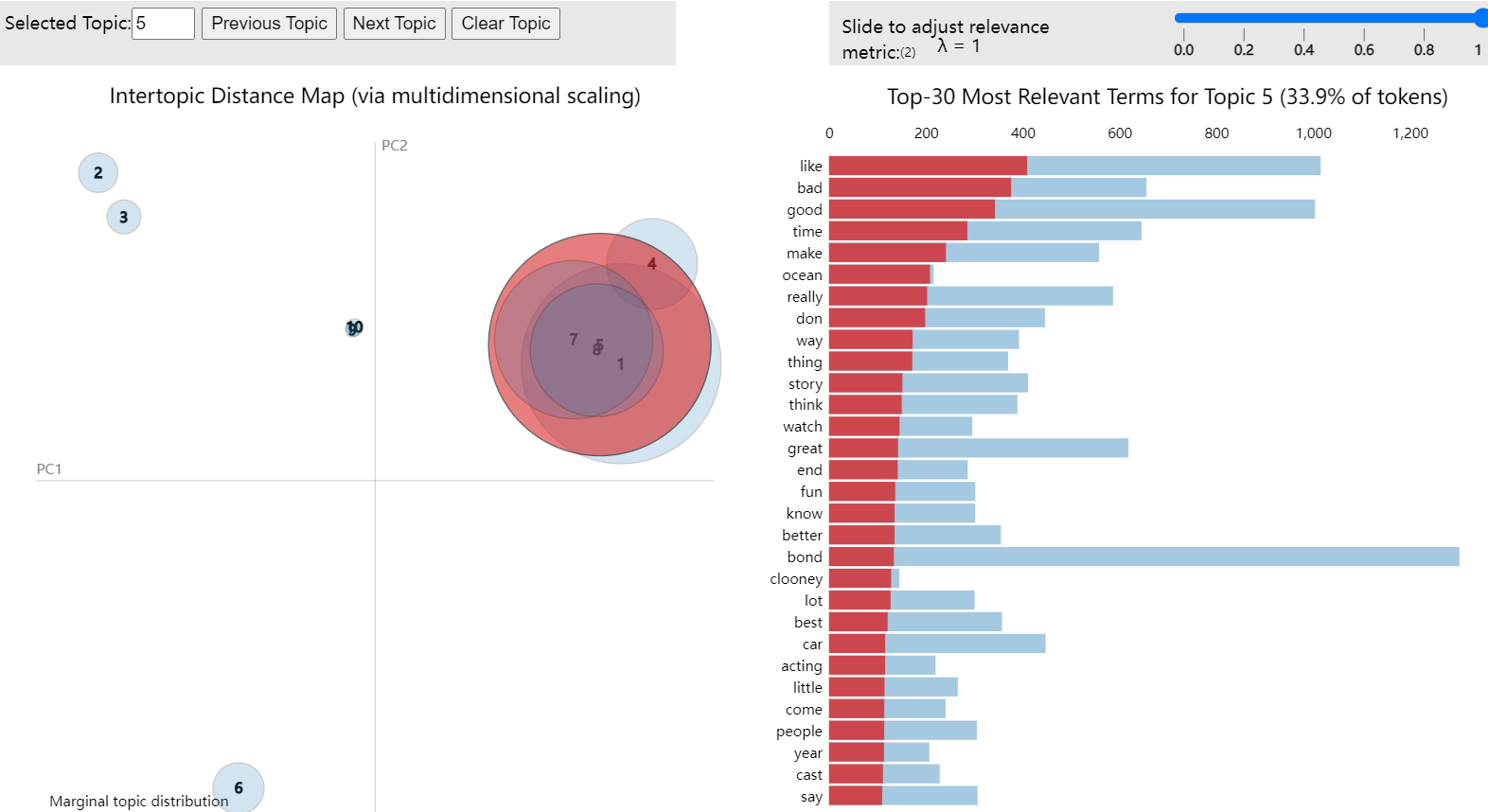
<https://doi.org/10.9781/ijimai.2018.12.005>

\

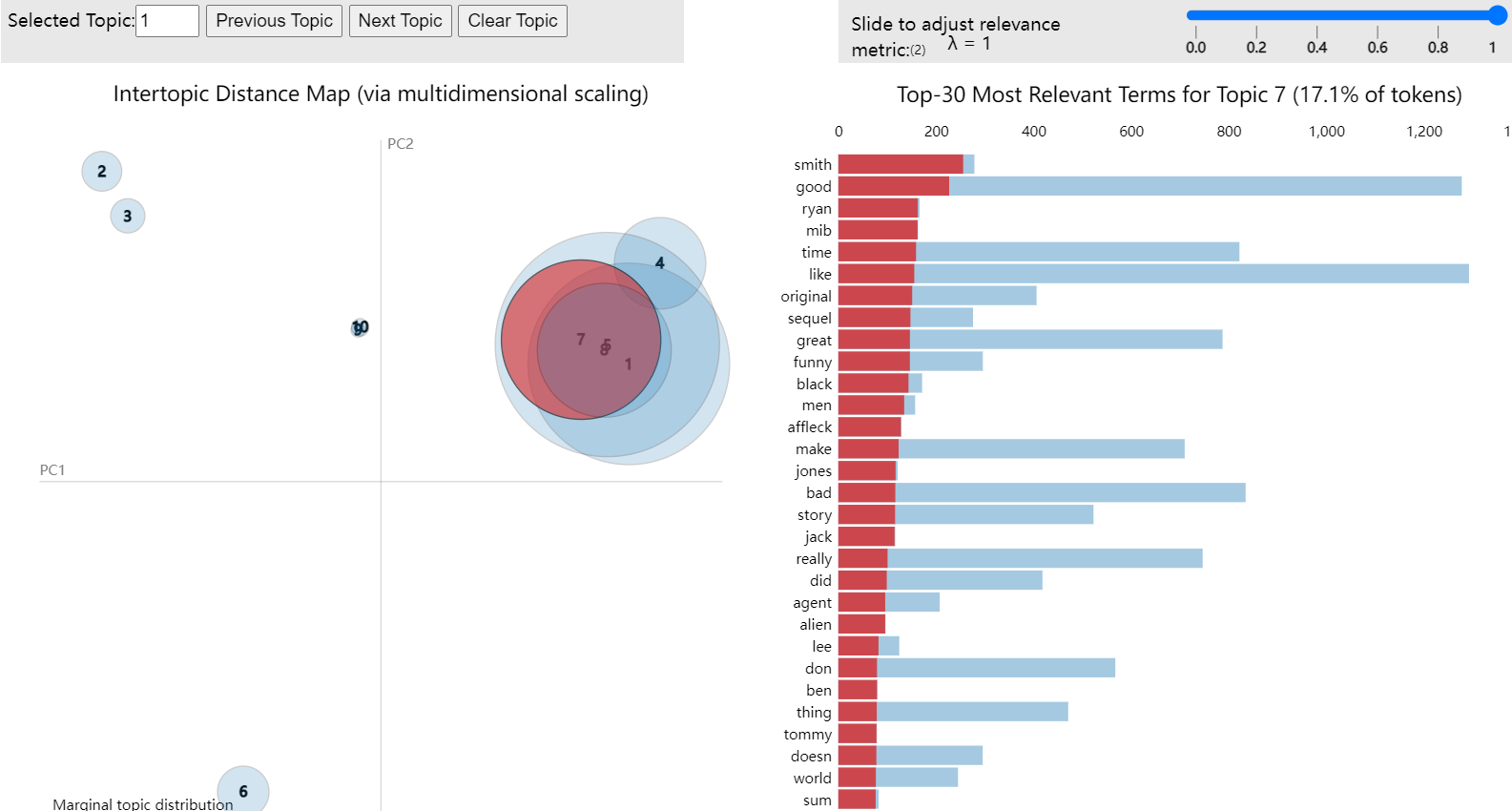
**Appendix**



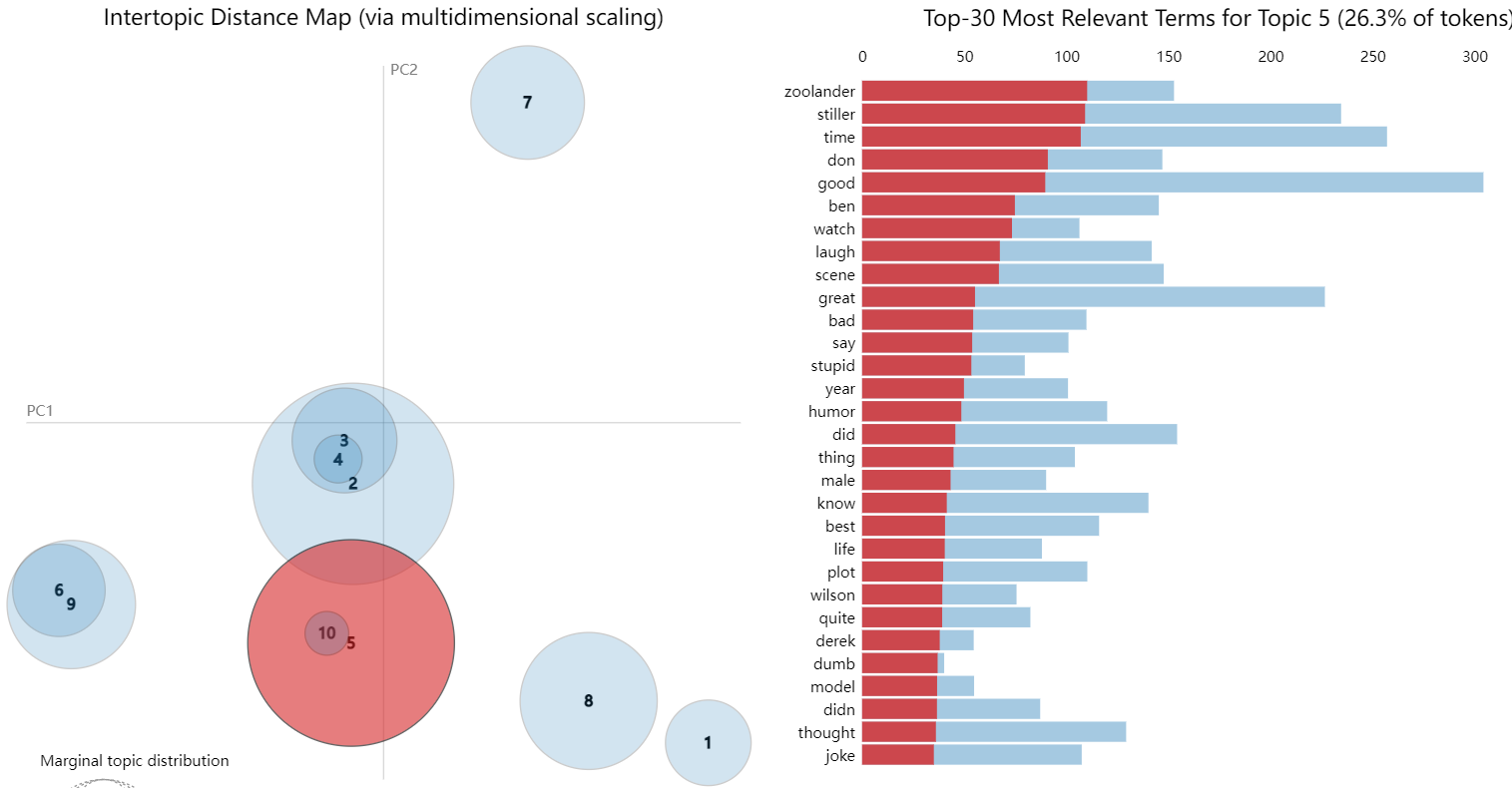
*Appendix 1. Word cloud of action movies (left) and comedy (reight)*



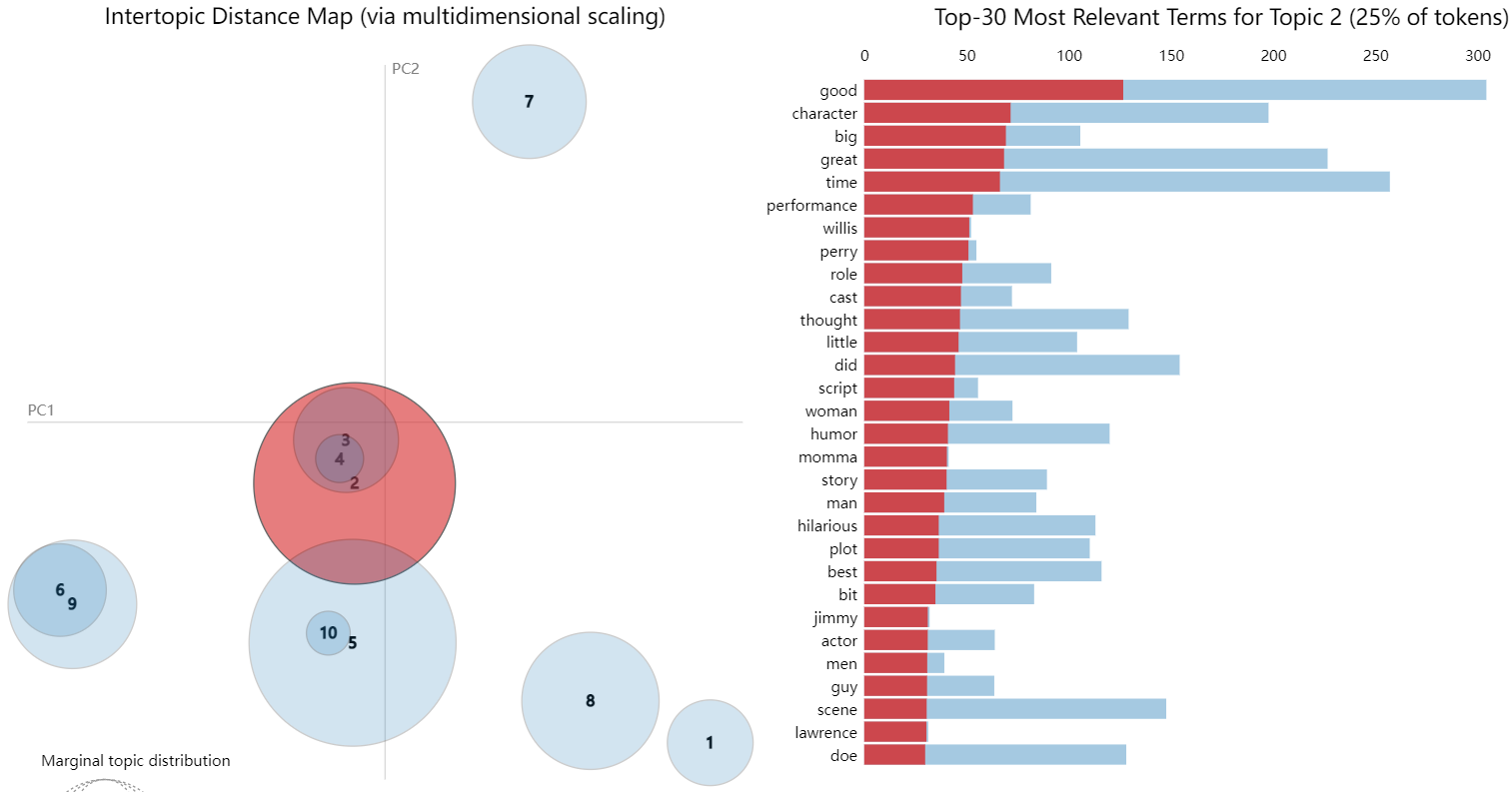
*Appendix 2. topic 2 of action movies*



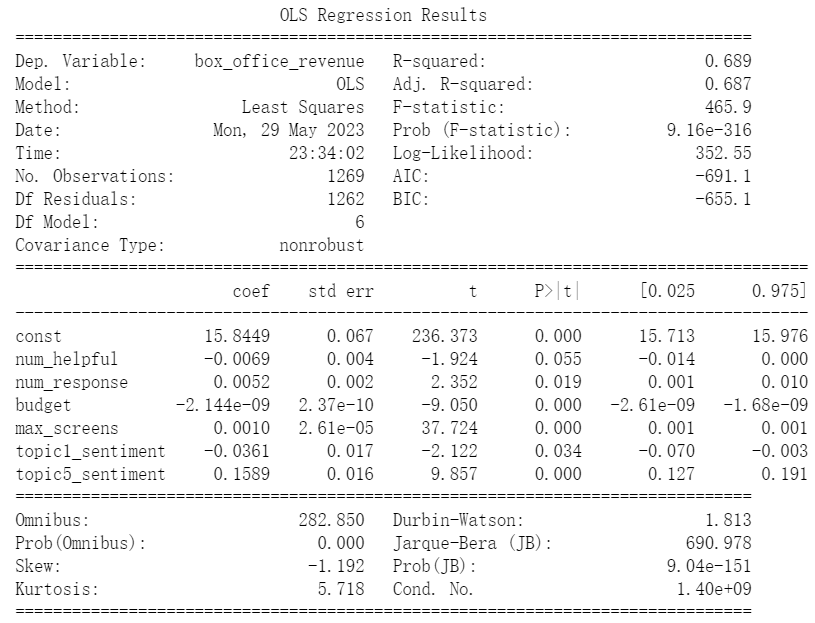
*Appendix 3. topic 1 of action movies*



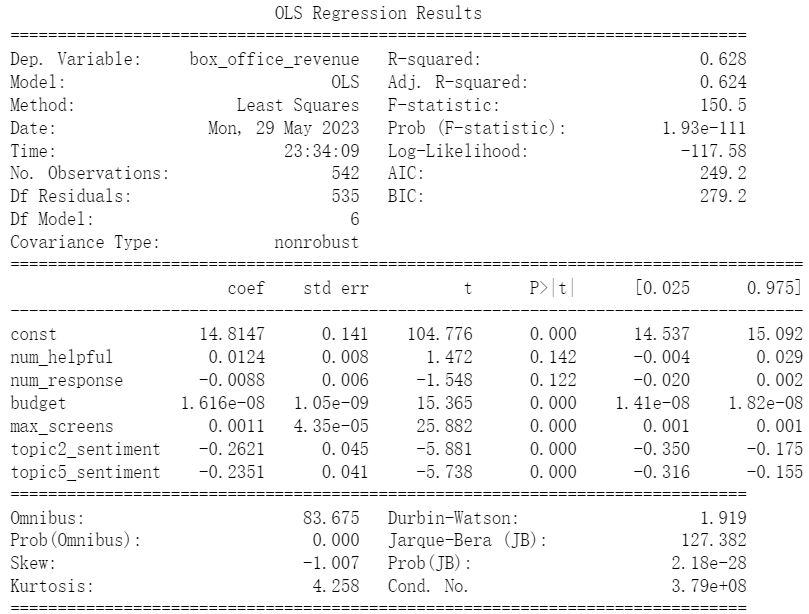
*Appendix 4. topic 5 of comedy movies*



*Appendix 5. topic 2 of comedy movies*



*Appendix 6. OLS regression results of action movies*



*Appendix 7. OLS regression results of comedy movies*