

Analysing and Tracking the Sentiment and Topics on Social Media

Table of Contents

1. Introduction	2
2. Problem Definition.....	2
3. Data Collection	2
4. Background & Data Exploration	3
5. Pre-processing	5
6. Method/Model	6
7. VADER Analysis	7
8. LDA Analysis.....	8
9. Conclusion	8
10. References	9

1. Introduction

In today's global market, companies can reach a diverse and large customer base. However, understanding customer satisfaction can be challenging due to the variety of feedback received for a product. This report aims to explore consumer opinions on the recently launched video game, Baldur's Gate 3, by employing social media text mining techniques and natural language processing models.

2. Problem Definition

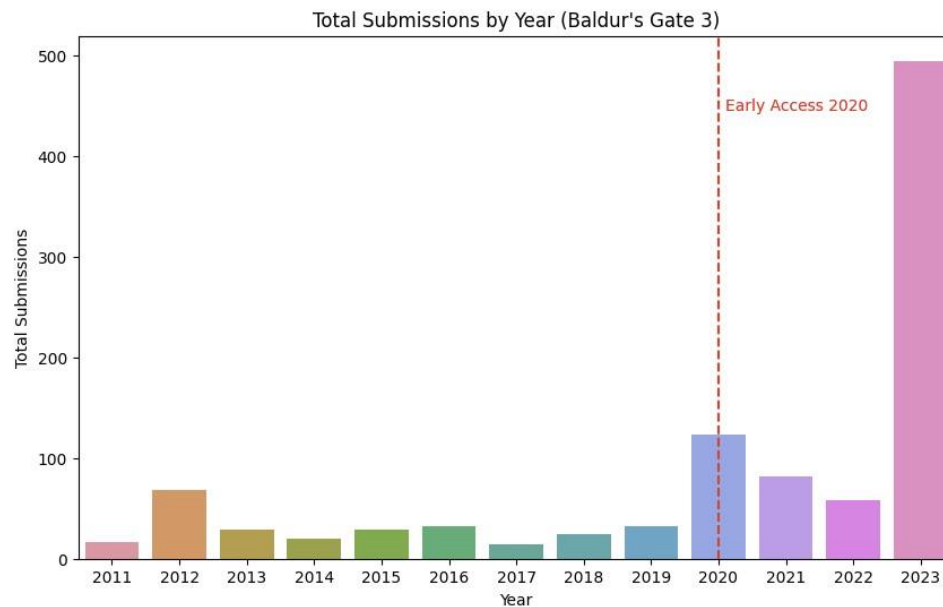
1. How has the sentiment towards Baldur's Gate 3 changed from development to release on the social media platform Reddit?
2. Which features are discussed the most with Baldur's Gate 3?

3. Data Collection

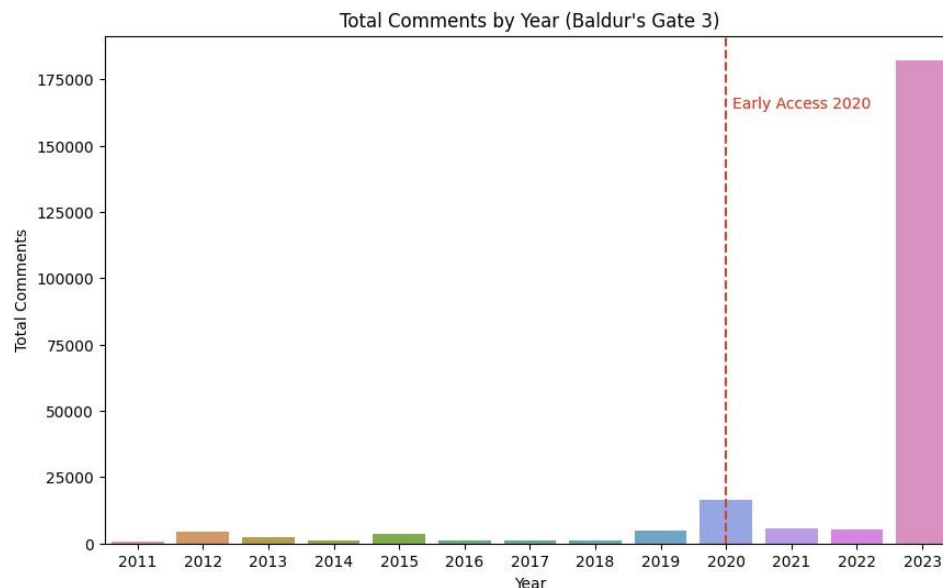
	Title	URL	Upvotes	Comment	SubmissionDate	CommentDate
0	Baldur's Gate 3 - Review Thread	https://www.reddit.com/r/Games/comments/15h6o61/baldurs_gate_3_review_thread/	2760	Updating the OP as the reviews are coming in. I'll also be updating it for a few days because I imagine most reviewers still didn't have enough time to finish the game. Full version of the game is already available on Steam.	2023-08-03 15:02:24	2023-08-03 15:04:00
1	Baldur's Gate 3 - Review Thread	https://www.reddit.com/r/Games/comments/15h6o61/baldurs_gate_3_review_thread/	2760	Review codes have only been out for a few days, so expect most platforms to still need a while to get their reviews out. A small price to pay for the game coming a month earlier.	2023-08-03 15:02:24	2023-08-03 15:06:05
2	Baldur's Gate 3 - Review Thread	https://www.reddit.com/r/Games/comments/15h6o61/baldurs_gate_3_review_thread/	2760	Take into consideration review copies got pushed pretty late (no doubt because of moving release date month earlier) so any outlet that puts score now probably either didn't finish it or rushed it to get on news wave. Or I guess played 20h/day. Either way see ya in Baldur's Gate, drop in for a pint.	2023-08-03 15:02:24	2023-08-03 15:08:56
3	Baldur's Gate 3 - Review Thread	https://www.reddit.com/r/Games/comments/15h6o61/baldurs_gate_3_review_thread/	2760	It leaked earlier but ACG said he spent 20 hours straight playing the game at one point because it was that good.	2023-08-03 15:02:24	2023-08-03 15:06:23
4	Baldur's Gate 3 - Review Thread	https://www.reddit.com/r/Games/comments/15h6o61/baldurs_gate_3_review_thread/	2760	As expected, sounds like Act 1 at the very least is of an immensely high quality - but I remember Divinity Original Sin 2's final hours were very rough at first and required a thorough overhaul when Larian was making the Definitive Edition. By the time it drops on console I'll have a good idea as to its quality in BG3's final act, but as long as it manages to be fun co-op with friends I'll likely buy in - and by all accounts, that's what the game has managed to nail. Very happy for Larian to have stuck the landing.	2023-08-03 15:02:24	2023-08-03 15:17:52

Using PRAW, the Python Reddit API Wrapper, data was collected over various subreddits, r/all, r/popular, r/gaming, r/pcgaming, r/gamingleaksandrumours, r/pcmasterrace, and r/truegaming with the term "Baldur's Gate 3" from June 11, 2011, to August 17, 2023, ensuring a wider variety of opinions as each subreddits has their own biases towards specific topics. Reddit's API limits the number of requests that can be executed. However, by using a "try-except" strategy along with a 60-second sleep interval, data collection was able to proceed uninterrupted for several hours. Each subreddit was collected separately, stored as a CSV and combined later. The dataset consisted of 229,981 rows, totaling 95MB and 133MB after processing. The CSV contains the submission title, URL shared, submission upvotes, every comment under the submission, the submission date, and comment date. Comments were stored with the same submission title, URL, upvotes and submission date.

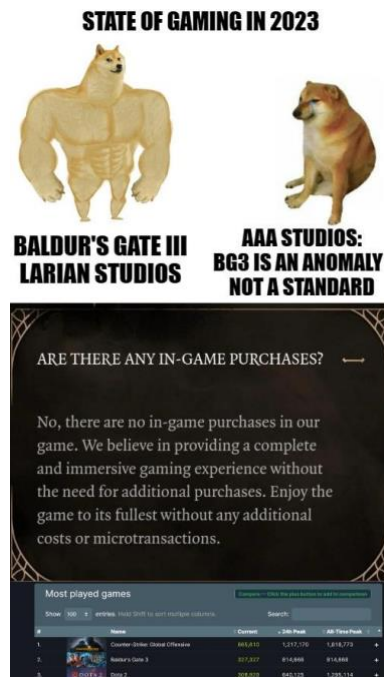
4. Background & Data Exploration



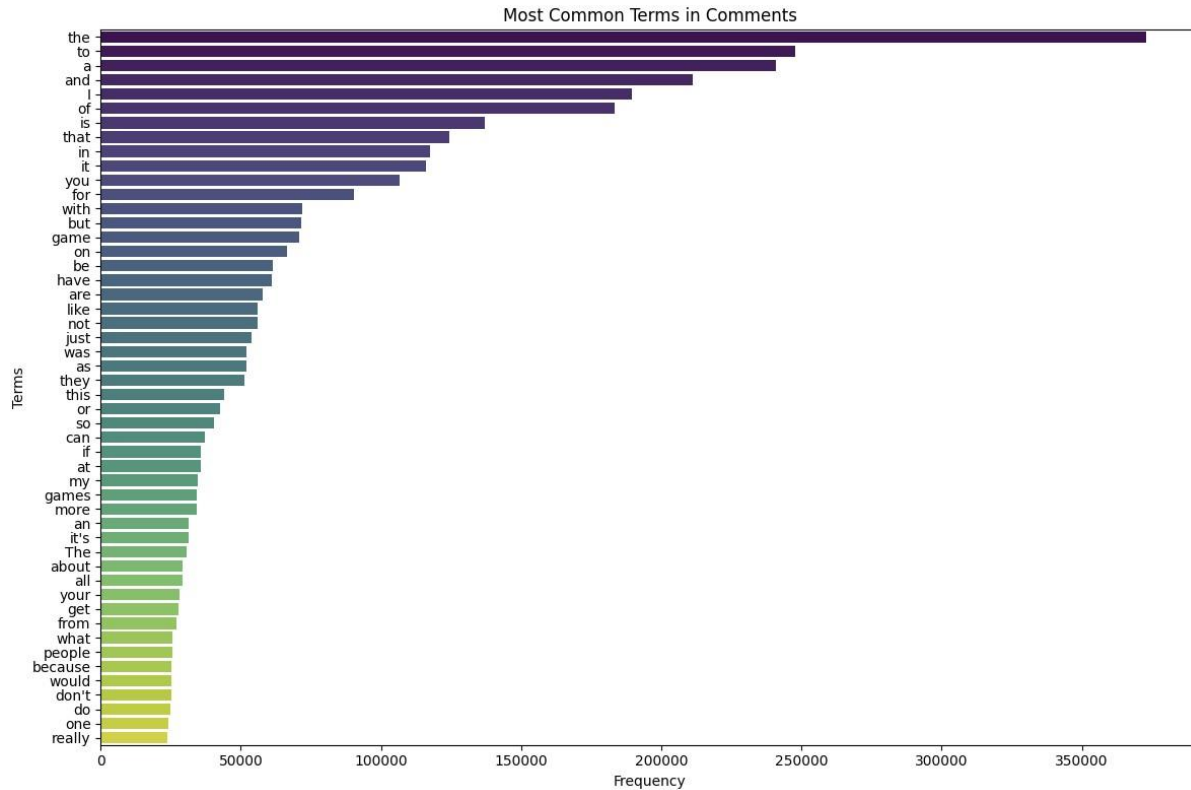
Baldur's Gate 3 was first developed by Black Isles, and rumors of its release started around 2011 and peaked the year after. Larian Studios acquired the intellectual property rights in 2017, and the first teaser was released in 2019. The early access started in 2020, generating a large buzz around the game. After releasing, the game became a critical success and is currently the highest rated PC game on Metacritic, and OpenCritic.



Submissions and comments follow a similar trend, but through comments, we are able to see the amount of activity revolving around Baldur's Gate 3. The submissions around the rumors did not get a lot of interactions compared to the early access and launch year.



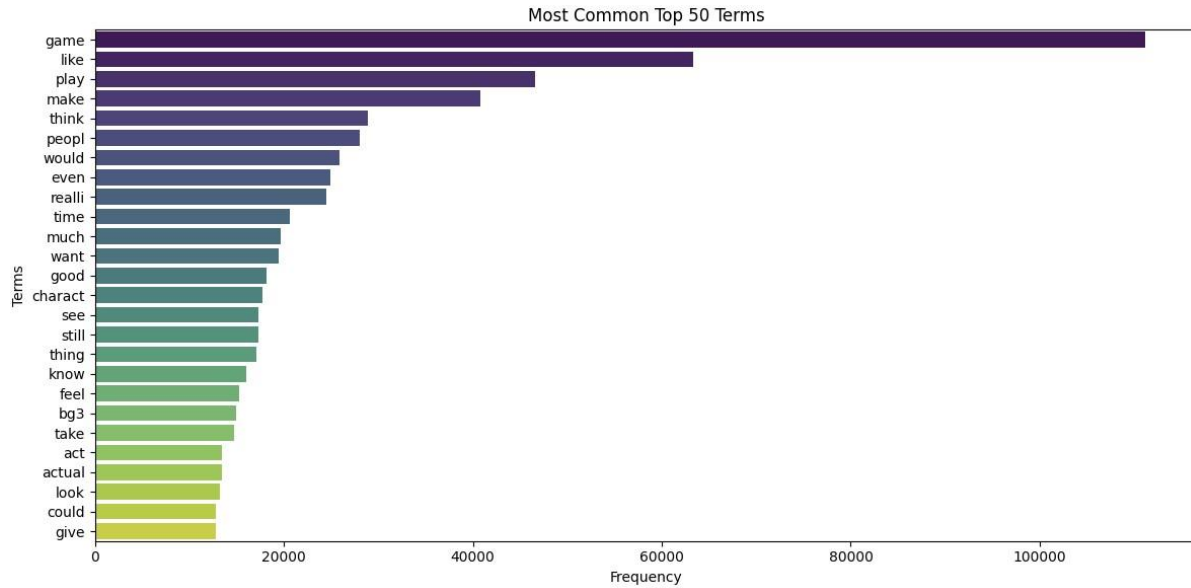
An analysis of the highest upvoted submission gives insights not merely into the game, but also about broader topics, such as how the gaming industry tailors game design for profitability. The top comments on this submission primarily focused not on Baldur's Gate 3 itself, but rather on the business practices of Larian Studio, specifically their decision to avoid incorporating microtransactions or additional fees for gameplay.



The top 50 words used around Baldur's Gate 3 do not show any useful insights without preprocessing. Compared to Twitter, Reddit allows users to write paragraphs of texts not allowing definite articles to be the most popular.

5. Pre-processing

To process the Reddit comments, we need to build a corpus by tokenizing the comment strings into a list of words, making sure each word is lowercase to control the variety and removing extra white spaces. Next, the text needs to go through a lemmatization process to break each word down to its root meaning. Stemming is then used to reduce an inflected word down to its word stem. Lemmatization and stemming words help reduce the variety of words with the same meaning such as "drawing," "draws" and "drawn" become "draw" for lemmatization and "player," "playing," and "plays" can be reduced to "play" as context. Stop words are then used to remove common grammatical words used in the English language such as "the," "this," and "it." Creating a custom stop word dictionary was needed to remove misspellings, internet slang, such as "game.", "/s," "oh" and "com."



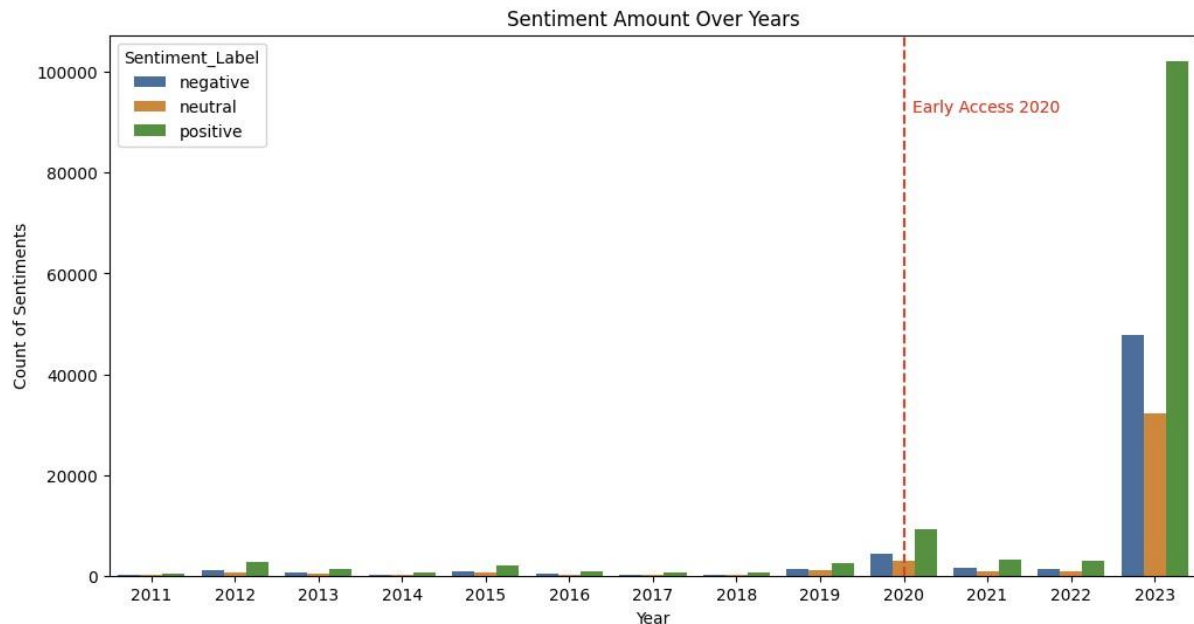
Data frame after preprocessing:

	Title	URL	Upvotes	Comment	SubmissionDate	CommentDate	Processed_Comment	Sentiment_Score	Sentiment_Label
0	Baldur's Gate 3 Proves That Pretty Much Everyone Still Wants Single-Player RPGs With No Microtransactions.	https://www.vg247.com/baldurs-gate-3-steam-launch-record-second-biggest-2023	19680	I don't think anyone wants micro transactions	2023-08-04 20:22:09	2023-08-04 20:22:09	think anyon want micro transact	0.0000	neutral
1	Baldur's Gate 3 Proves That Pretty Much Everyone Still Wants Single-Player RPGs With No Microtransactions.	https://www.vg247.com/baldurs-gate-3-steam-launch-record-second-biggest-2023	19680	Co op is a big part for me	2023-08-04 20:08:17	2023-08-04 20:08:17	co op big part	0.0000	neutral
2	Baldur's Gate 3 Proves That Pretty Much Everyone Still Wants Single-Player RPGs With No Microtransactions.	https://www.vg247.com/baldurs-gate-3-steam-launch-record-second-biggest-2023	19680	[removed]	2023-08-04 20:55:22	2023-08-04 20:55:22	[removed]	0.0000	neutral
3	Baldur's Gate 3 Proves That Pretty Much Everyone Still Wants Single-Player RPGs With No Microtransactions.	https://www.vg247.com/baldurs-gate-3-steam-launch-record-second-biggest-2023	19680	This was something that absolutely didn't need proving. What a silly title	2023-08-04 20:36:36	2023-08-04 20:36:36	someth absolut need proving. silli titl	0.0258	positive
4	Baldur's Gate 3 Proves That Pretty Much Everyone Still Wants Single-Player RPGs With No Microtransactions.	https://www.vg247.com/baldurs-gate-3-steam-launch-record-second-biggest-2023	19680	I want multiplayer rpgs with no micro transactions.	2023-08-04 19:53:47	2023-08-04 19:53:47	want multiplayer rpg micro transactions.	-0.2263	negative

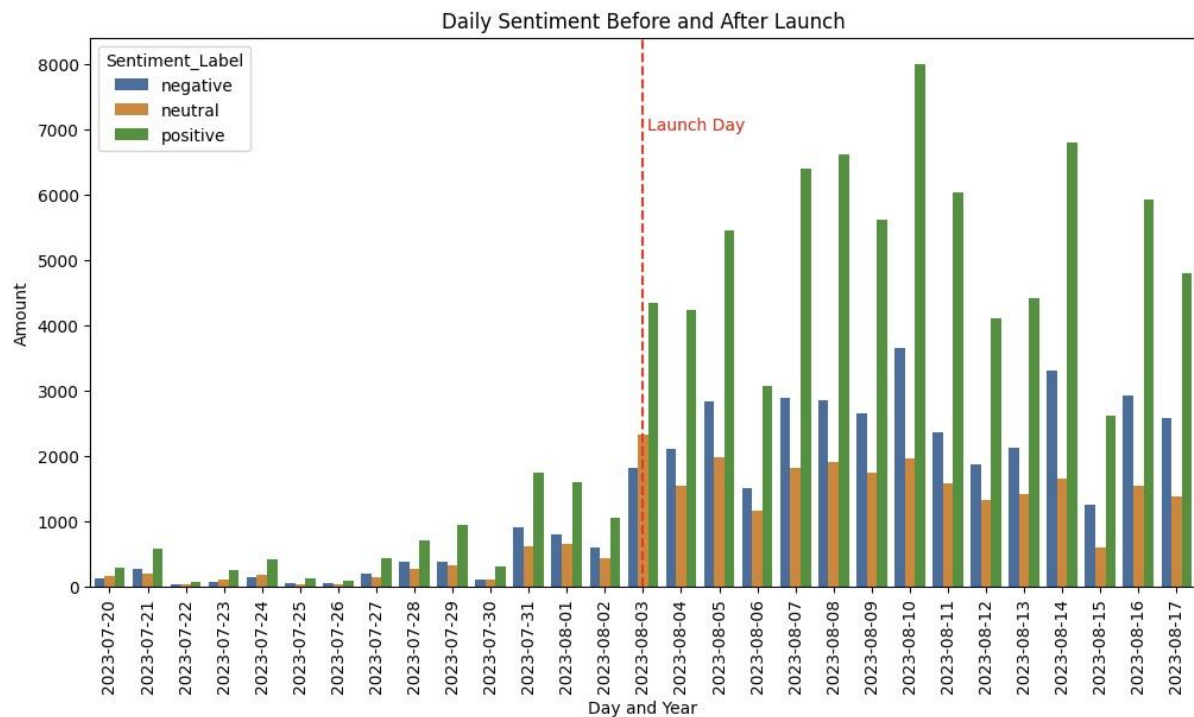
6. Method/Model

To answer how the sentiment towards Baldur's Gate 3 changed from development to release on the social media platform Reddit, I chose the VADER (Valence Aware Dictionary for sEntiment Reasoning) model to guage sentiments of each comment, and LDA (Latent Dirichlet Allocation) for topic modelling to find out which features are discussed the most with Baldur's Gate 3. VADER was chosen because it was built with social media in mind and has a more sophisticated approach to weighing sentiments (emojicons, internet slang "lol," exclamation marks, etc.), but the lexicon does not seem to be continuously updated. Although LDA is slower than LSA (Latent semantic analysis), the goal is to find the underlying topics from all the comments.

7. VADER Analysis



Baldur's Gate 3 has had an overall positive sentiment since the very beginning. It was met with high praise during the early access launch in 2020 and the sentiment during the launch year was amplified. Although the negative sentiments grew much higher than neutral from 2020-2023, it could be attributed towards business practices in the gaming industry.



Leading up to the launch date (2023-08-03), there was an overall positive sentiment for Baldur's Gate 3, and less positive sentiment the day before launch. During the launch day, the number of positive comments exponentially increased, indicating a successful launch. A rare situation also occurred as most review sites did not release their reviews before or on the launch day as they had not completed the game.

8. LDA Analysis

[illegible]

Baldur's Gate 3 is a character role playing game that follows the Dungeons & Dragons 5th edition rule system. An LDA analysis revealed four primary themes that contribute to the game's characteristics. The first topic examines the success of the game's launch day, contrasting it with Fallout games that typically debut with numerous bugs and problems. The second topic is centered on gameplay, specifically character creation and the unique personas you can encounter. Topic three discusses in-game mechanics, including the types of weapons available, skills used, and the grading of items and abilities. The final topic captures positive opinions about the gaming experience.

9. Conclusion

Based on the two different methods of sentiment analysis, Baldur's Gate 3 has received positive reactions. The main discussion points among players were the enjoyable gameplay mechanics, compelling characters, and a problem-free launch. Highlighting these aspects can provide stakeholders with insights into what makes a game successful and what players appreciate most. Usually, popular games debut with immediate review scores which could introduce bias into the data, but Baldur's Gate 3 did not have reviews during their launch day. The ability to correlate player sentiment with professional reviews suggests that these models are effective in pinpointing key features of a product and capturing the overall mood on social media.

10. References

- [1] Metacritic. (2023). 'Top PC Games by Metascore,' Metacritic. Available at: <https://www.metacritic.com/game/pc> (Accessed: 30/08/2023).
- [2] OpenCritic. (2023). 'Top PC Games by Score,' OpenCritic. Available at: <https://opencritic.com/browse/pc> (Accessed: 30/08/2023).
- [3] Pykes, K 2023, 'Stemming and Lemmatization in Python', DataCamp, viewed 28 August 2023, <https://www.datacamp.com/tutorial/stemming-lemmatization-python>.
- [4] Calderon, P. 2017, 'VADER Sentiment Analysis Explained', Medium, viewed 30 August 2023, <https://medium.com/@piocalderon/vader-sentiment-analysis-explained-f1c4f9101cd9>.