A Case Study of 'Hogwarts Legacy' Reviews Through Social Media Analysis

1.Introduction

In the ever-evolving field of digital entertainment, video games not only provide a way to enter immersive worlds, but also spark vibrant discussions and interactions between gamers. Steam is a digital game distribution service developed in 2003 by game development company Valve Corporation that allows developers to distribute their games to millions of users. Currently, the Steam catalog contains approximately 30,000 games and accounts for more than half of the global market share of computer game sales. Service statistics report an average of 18 million active Steam users（Bounie et al., 2005）. Similar to the Apple App Store, an online game distribution platform such as Steam also provides users with the ability to write reviews of all their games and recommend (or not) games to other players, so Steam can be considered a social platform for gamers. One of the popular games on the Steam platform, Hogwarts Legacy, provides a unique opportunity to explore these interactions. Set in the Harry Potter setting, this game appeals to a wide audience and is an ideal candidate for social media analytics research. Previous work on mobile app reviews has shown the value of studying reviews (Khalid et al. 2015; Pagano and Maalej 2013). This analysis focuses on the social media aspects of the game, specifically through user comments on Steam. By examining these comments, this report aims to reveal insights into player sentiment, preferences and engagement patterns.

This paper applies web and social media analytics techniques to extract, analyze and interpret data from Steam. This will not only enhance our understanding of the popularity of games, but also contribute to the wider market research strategy of the gaming industry. Through this analysis, we aim to provide actionable insights that will inform Hogwarts Legacy game developers, marketers and Steam community managers about current trends and player satisfaction.

In the first part of this paper, we compare web and social media data analysis with the traditional methods of consumer and marketing research in the literature review section, comparing the differences and connections. Next, we analyze the selection of methods and tools and explain why sentiment analysis, word cloud, K-means clustering and LDA topic modeling were chosen. In the second section, the methodology of this paper will be described, including data collection, data preprocessing and data analysis. The analysis methods include sentiment analysis, word cloud and LDA topic modeling. In the third part, the paper discusses the analysis results obtained. In the fourth part, recommendations for game developers, Steam platform and game marketers will be given. In the fifth part, the effectiveness and limitations of the analytic methods used will be critically evaluated.

2.Literature Review

In recent years, social media has significantly shaped marketing strategies. Marketers are experiencing positive outcomes by utilizing the tools and techniques available on these online platforms. With the introduction of the Internet and the development of technology, management models and marketing strategies have also changed. (Cicero, 2014)

Traditional consumer and marketing research methods cover a wide range of techniques and strategies designed to collect and analyze data on consumer behavior and market dynamics. These methods mainly include:

Survey research: Consumer opinions and preferences are collected through questionnaires. These questionnaires can be conducted face-to-face, over the phone, by mail, or online.

- Focus Groups: Gathering a small group of target audiences for a moderated discussion to gain insight into their feelings and opinions about a specific product, service or advertisement.
- In-depth interviews: Detailed one-on-one conversations with individual consumers to gain deeper insights.
- Behavioral Observation: Observe consumers' actual behavior in a natural environment rather than relying on their self-reporting to obtain more realistic behavioral data.

Social media analytics is the practice of collecting data from social media platforms and analyzing the data to help decision makers address specific issues. Social media analytics has been widely used by groups such as social science researchers, senior corporate decision makers and healthcare professionals. The following are common methods of networking and social media:

- Data Acquisition: automated tracking and collection of large amounts of user-generated content and interaction data through social media platforms.
- Sentiment Analysis: Use natural language processing techniques to analyze social media texts for sentiment tendencies such as positive, negative or neutral.
- Trend Identification: Identify market trends and changes in consumer interests by analyzing topics, hashtags, and user behavior.

There has been quite a bit of literature that has compared this traditional consumer marketing research methodology with the social media marketing methodology.Businesses increasingly prefer social media marketing over traditional methods due to significant advantages in terms of time, audience reach, collaboration, and cost (Holmes, 2017). Consumer behavior analysis process is now easier because of social media (Durmaz & Efendioglu, 2016). In comparison to traditional marketing, social media marketing strategy enables the company to target its potential consumer in a large number of audiences, with a very low financial cost (Mangold & Faulds, 2009). Moreover, product advertising and promotion, through social media marketing compared to traditional one (e.g. through TV advertising), can be done at a very low cost by achieving the same level of impact and awareness in relation to potential customers (Hainla, 2018). A study conducted in Kosovo indicates that 93.3% of companies were optimistic about the digitalization of marketing, while 14.3% declared that digitalization would bring lower expenses for marketing activities compared to other forms of marketing and channels (traditional) (Kempt, 2018). When facing the application of social media analytics (SMA), there are a series of challenges, including handling unstructured and free-form content, ensuring the authenticity and accuracy of data, dealing with the real-time and dynamic nature of data, managing the diversity and large volume of data, and developing complex analytical methods (Holsapple et al., 2014). But overall, social media analysis provides a faster and cheaper way to understand and predict consumer behavior than traditional market research tools, enabling organizations to respond more flexibly to market changes and effectively improve business strategy and operational efficiency.

Advances in digital technology continue to change purchasing behavior, and in an environment where customers expect brands to deliver a more personalized experience, companies need to provide customer support in places they never have before.In this case, improving customer service on social media platforms is crucial for brand manufacturers for a variety of reasons. Firstly, the company urgently needs to achieve a higher level of customer satisfaction. Due to the historical low level of consumer trust in businesses in 2018, brands may lose a significant number of customers who are dissatisfied with customer service (Pasternak, 2017). From a positive perspective, improving customer satisfaction can help increase revenue and enhance customer loyalty. The popularity of online platforms has allowed for more immediate interactions between brands and their customers, with brands being able to respond to customer feedback in a timely manner and provide better customer service. With a favorable buying experience, customers can even repurchase from different channels, leading to increased profits and customer loyalty (eMarketer, 2017).Additionally, customers who experience a brand's convenient customer service may recommend that brand to others, which is the purpose of word-of-mouth marketing. According to Nielsen (2012), 92% of internet participants say they trust recommendations from friends and family more than any other type of advertising, so better customer service could further lead to more profits. Therefore, customer satisfaction with customer service can be considered as the KPl of a company.

To quantify this KPl, this study applies sentiment analysis to customer feedback.The purpose of sentiment analysis is to extract information from text and categorize sentiment into positive, neutral, and negative categories based on polarity scores (Pozzi et al., 2017). By converting text into scores, sentiment analysis allows for better visualization of customer satisfaction and helps to examine the causes of negative customer feedback. Therefore, emotional analysis is an ideal method for quantification.

3.Selection of analytical methods and tools

3.1 Sentiment Analysis

Sentiment analysis, also known as opinion mining, is a key process for understanding the polarity of a text in context and determining whether the text is positive, negative or neutral. This method helps to extract and analyze the opinions or attitudes expressed by individuals (Devika et al., 2016). This method is particularly important for understanding the public's overall feelings about a brand or product.It helps marketers to assess the impact of marketing campaigns and consumer reactions to new products or events. This article uses the RoBERTa model from sentient analysis.RoBERTa (Robustly optimized BERT approach) is a pre trained model for natural language processing (NLP) developed by the Facebook AI research team. RoBERTa has achieved excellent results in multiple benchmark tests of NLP, such as GLUE, SQuAD, and RACE.

3.2 Word Cloud

Word cloud is a graphical representation method that visually displays the most frequently occurring vocabulary in text data. Usually, word clouds are used to summarize text and assist in text analysis by displaying key themes without providing context related to language connections or meanings. This method is mainly statistical and provides minimal interaction functionality (Heimerl et al., 2014). Through word clouds, analysts can quickly identify hot topics and key terms in social media discussions, thereby better understanding the public's concerns and interests.

Fig. 1. Word Cloud Example

3.3 LDA Topic Modeling

Latent Dirichlet Allocation (LDA) is a topic model that can identify different topics from a large amount of text data. LDA can be used to aggregate, cluster, and link very large data, as it generates a weighted list of topics for each document (Hindle&Campbell, 2014). Applying LDA topic modeling to social media data can reveal the main themes of user discussions, enabling marketers to capture market trends and adjust their content strategies accordingly.

4.The Case: Hogwarts Legacy game

4.1 Data Collection and Pre-processing

4.1.1Data collection

For my project, I needed to gather user reviews of the video game "Hogwarts Legacy" to analyze consumer sentiment and identify common themes in player feedback. The dataset studied in this paper is from Kaggle.The data was extracted using the Steam API's 'GetReview' endpoint, which allows for the retrieval of user reviews based on various parameters. The authors of the dataset specified the application ID of 'Hogwarts Legacy' and set the parameters to fetch a total of 5040 game reviews of the game on the Steam platform.

4.1.2 Data Cleaning

Remove Duplicates: Ensure that there are no duplicate entries in your dataset. If reviews are collected multiple times, they should be identified and removed.

Handle Missing Data: Check for missing values in critical fields such as the review text or ratings. Depending on the analysis, you might choose to either fill the missing values with a placeholder, average, or median value, or you might decide to drop these entries.

4.1.3 Text preprocessing

First, the required libraries were installed and imported, including Textblob and NLTK, and English stop words were downloaded. Next, a function was defined to clean up the text of deactivated words while removing the predefined list of deactivated words as well as URL links, punctuation marks and numbers. Then, the text was further processed using stemming extraction and word form reduction. Finally, the process of word form reduction was completed.

4.2 Sentiment Analysis

In this paper, a pre-trained RoBERTa model is used to perform sentiment analysis on the collected comment data.First, the script loads the RoBERTa model optimized for sentiment analysis and its corresponding splitter for processing the text data into a format acceptable to the model. Next, a function sentiment_scores is defined to process comment batches, which transforms the text batches into the input format needed by the model and uses the model for forward propagation to obtain sentiment scores. Another function, sentiment_category, then categorizes the reviews as 'NEGATIVE', 'NEUTRAL', or 'POSITIVE' based on the score. This analysis helps to understand the overall feeling of gamers towards the game, which can be used to guide product iteration, improve user experience, and develop marketing strategies. The specific results are shown in Figure 2, where it can be seen that most of the reviews are positive, which indicates a high level of satisfaction among game consumers.

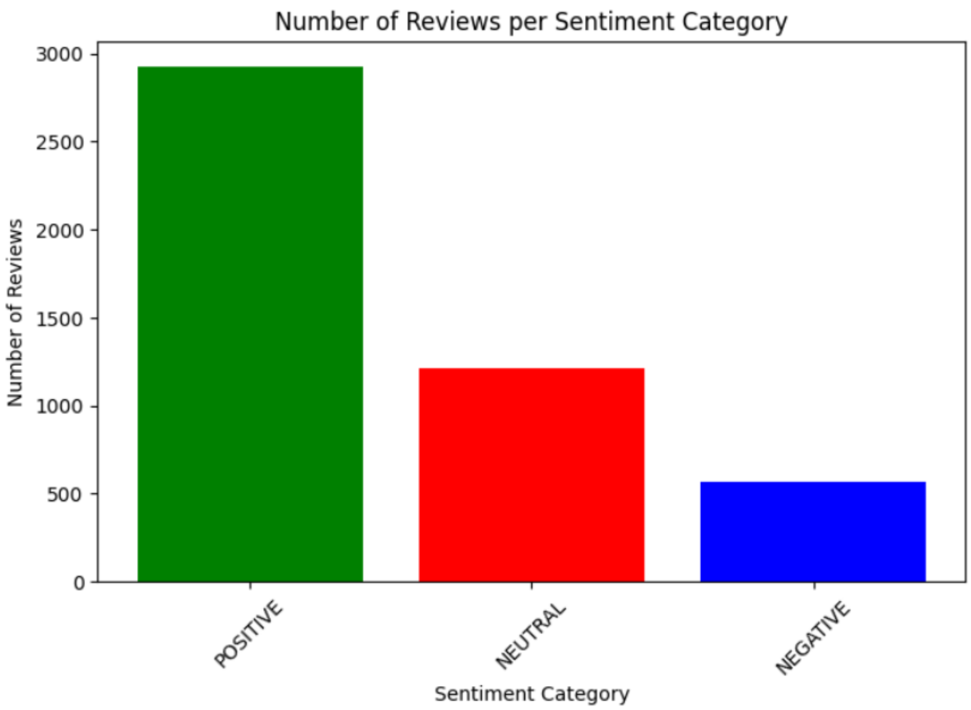

Fig. 2. Sentiment Category

4.3 Word Cloud

Before introducing clustering and topic modelling approaches, a word cloud can be generated to roughly overview keywords in negative tweets.

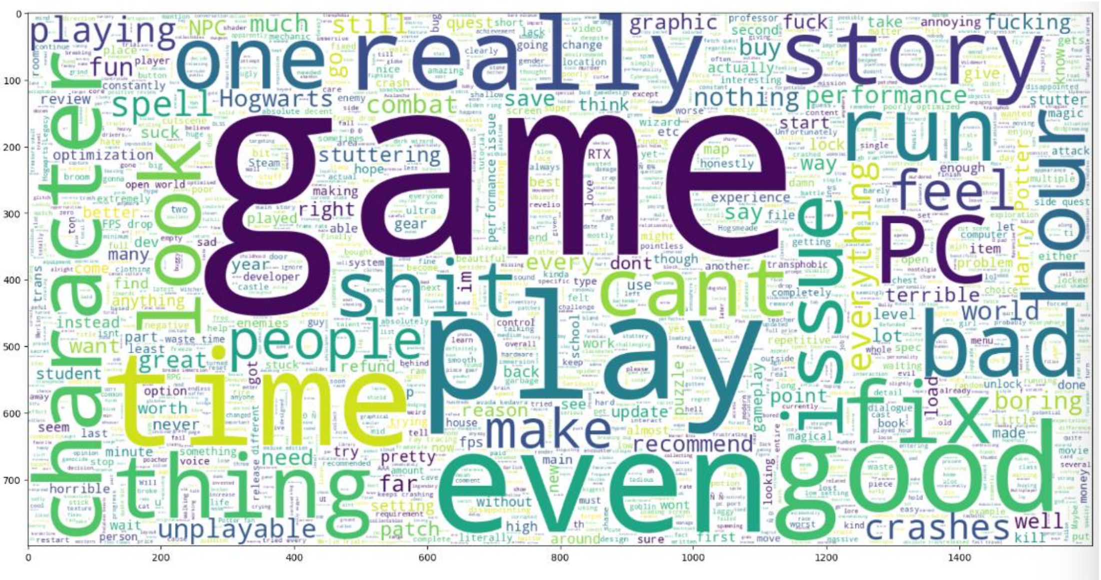

Fig. 3. Word Cloud for Text Generation with Negative Emotional Comments

As can be seen in Figure 3, the Word Cloud reflects people's comments on the game Hogwarts Legacy. This word cloud shows a variety of comments from players about the game Hogwarts Legacy. Prominent words in the word cloud include “game,” “story,” “play,” and “good,” which suggests that the comments were made about the game. “This suggests that the game's story and gameplay are frequently mentioned in the reviews and have positive emotional overtones. However, there were also negative words such as “bad”, “shit”, “unplayable” and “crashes However, there are also negative words such as “bad”, “shit”, “unplayable”, and “crashes”, suggesting challenges and dissatisfaction with performance issues, game crashes, and playability. Words such as “spell,” “Hogwarts,” and “combat” reflect the game's content's close association with the Harry Potter series, and the players' discussion of the game's magical elements and combat system. Overall, the word cloud demonstrates the diversity of feelings and the wide range of feedback from players about the game experience, storyline, and performance.

4.4 LDA Topic Modeling

The LDA method (Blei, 2012) is a powerful technique for revealing unobserved word groups in textual data (e.g., customer reviews).As an unsupervised machine learning algorithm, LDA can analyze and cluster large text databases to discover hidden patterns or themes by measuring the probability of observing words in similar word clusters (Reisenbichler and

Reutterer, 2019).This paper presupposes that this model requires a pre-set number of 10 themes.After running the LDA model, the distribution was plotted as 10 graphs representing six different groups.

In Figure 4, the x-axis of each graph represents the topic terms, while the y-axis indicates the weight of these terms in a given topic.A high bar means that the relevant term is more relevant to the topic it belongs to.Figure 4 presents the results of the LDA topic model analysis, which shows the distribution of keywords for 10 different topics.These keywords are the 10 most highly weighted words in each theme weight.From the figure, it can be seen that these themes may be related to the Harry Potter game or related content because “Harry”, “Potter”, “Harry”, “Potter”, and “game” appear in several themes.

Each theme consists of a series of words that represent the central idea or content of the theme. Themes 0, which includes the words “Harry” and “Potter”, may be part of a general discussion of the Harry Potter games themselves, and may refer to the overall experience or concept of the games. It may refer to the overall experience or concept of the game. These terms are not entirely visible in Theme 1, but the presence of the word “game” and a few other ambiguous words may indicate a theme related to game mechanics or functionality. Theme 2 contains combinations of capitalized letters, including “GAME” and “LOVE,” and possibly spellings such as “Avada Kedavra,” with a focus on emotional responses to the game and specific game elements. Emotional responses and specific game elements such as spell casting. These words are not fully visible in Theme 3, but appear to have names (possibly spells or characters) and possibly some game elements. Terms such as 'game' and 'world' in Theme 4 suggest discussion around the context or environment of the game, and 'combat' denotes the action elements of the game. In Theme 5, terms such as 'gameplay', 'good', 'story', and 'amazing' indicate positive feedback about the narrative and enjoyment of the game. fun. Topic 6 seems to focus on the experiential part of the game, with terms like “awesome” and “Hogwarts” suggesting an emphasis on the immersive aspects of the game world. Topic 7 seemed to focus on narrative elements, such as “Harry,” “story,” and “great,” indicating discussion around plot and characters. Theme 8, which includes words such as 'game', 'play', 'bad' and 'good', expresses a variety of views and can represent an opinion on the possibility of play. This theme includes words such as “game”, “play”, “bad” and “good”, showing a wide range of opinions and representing general feedback on the playability and quality of the game. The terms “played”, “best”, and “game” in Theme 9 indicate reflections on the overall game experience, possibly comparing it to other games or discussing replay value.

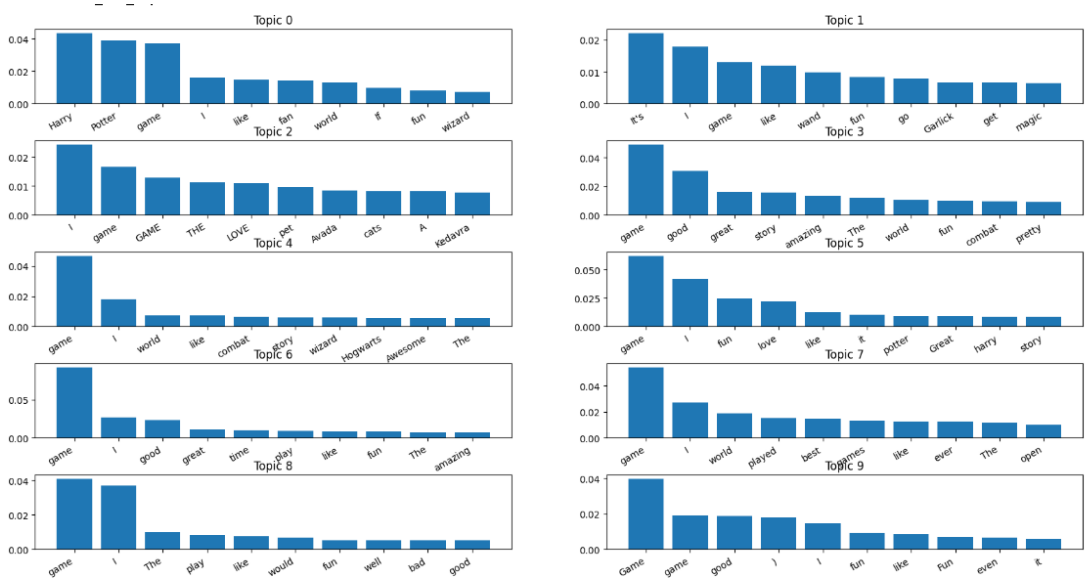

Fig. 4. The topic distribution of LDA

5. Business Implications

Based on the conclusions drawn from the previous analysis and existing literature, the following are recommendations for Hogwarts Legacy game developers： Avalanche Software

Avalanche Software is an electronic development studio that is currently a subsidiary of Warner Bros. Interactive Entertainment. In addition, previous research has shown that gamers are an extremely difficult user group to satisfy (Chambers et al., 2005), which makes the quality of games an important issue. In order to improve the perceived user quality of games, a better understanding of gamers' concerns is crucial for game developers. Combined with the previous analysis, this paper gives the following recommendations:

- Stability and performance of the game should be improved. Regular updates and patches that address bugs and crashes are essential to maintaining a positive player experience.
- Forums and social media can be utilized to communicate upcoming fixes and updates. Player feedback can be a goldmine for improving the game and should be actively sought and recognized.
- Game development teams should continue to develop and enhance player-favorite features such as immersive world-building and compelling storytelling. These elements are often

emphasized in positive reviews.

- Timely handling of negative feedback. Service quality is the most important factor influencing online game satisfaction and has the greatest total impact on online loyalty (Yang et al., 2009). Based on the previous analysis of game negativity, game developers can explore the root causes of players' challenges and dissatisfaction with performance issues, game crashes, and playability to improve overall service quality.
- Enhance community functionality by creating dedicated discussion spaces in Steam for topics identified in the LDA modeling (e.g., game mechanics and story elements) to increase user engagement.
- Improve the targeting of game content marketing. Create content that is consistent with the content that players are most interested in using the keywords and phrases identified in LDA topic modeling. For example, blog articles or videos about game strategies or legends can attract potential game consumers.

6. Limitations and Future Research Directions

Although the methods used in this social media analysis, including sentiment analysis, word cloud visualization, and LDA topic modeling, provide insightful insights into player feedback and game content discussions, there are still some limitations that need to be addressed in future research. Firstly, the amount of data is insufficient, as it is difficult to obtain a large amount of data within a limited time due to the limitations of the Steam API. Future research can expand the scope of data collection, including more platforms and a wider time frame, to capture players' more comprehensive emotions and perspectives. Secondly, during a certain period of time, if the game has discounts and promotions, it may lead to biased review data, which can result in bias in the review dataset. In addition, although sentiment analysis and word clouds provide quantitative observations of data, they may simplify the complexity and subtle differences of user emotions. Further qualitative analysis is required through manual review or advanced natural language processing techniques to fully understand the context and nuances of user feedback. In the future, employing more advanced machine learning models that can better handle the nuances of human language, such as transformer based models that focus on contextual nuances in text, can enhance the accuracy and depth of sensitive analysis. This study mainly focuses on English content and may not fully capture the global perspective of popular franchising like Harry Potter. Incorporating multilingual and cross-cultural analysis can provide a more global understanding of player emotions.

References

Bounie, D., Bourreau, M., Gensollen, M. & Waelbroeck, P., 2005. The effect of online customer reviews on purchasing decisions: the case of video games.

Khalid H, Shihab E, Nagappan M, Hassan AE (2015) What do mobile app users complain about? IEEE Softw 32(3):70–77

Pagano D, Maalej W (2013) User feedback in the appstore: An empirical study. In: 21St international requirements engineering conference (RE). IEEE, pp 125–134

Cicero, N., 2014. Marketing Statistics Every CMO Should Know. Available from: http://www.socialfresh.com/marketing-statisticsevery-cmo-should-know-in2014/?utm_medium=Webbiquity.com [Accessed 4 May 2024].

Holmes, S., 2017. Social Media Marketing 2018: How to boost your company's success with Facebook, Twitter, Instagram & Co. Kindle Edition. The Open Web Learning Institute.

Durmaz, Y. & Efendioglu, H.I., 2016. Travel from Traditional Marketing to Digital Marketing. Global Journal of Management and Business Research: E Marketing, 16(2).

Mangold, G.W. & Faulds, D.J., 2009. Social media: The new hybrid element of the promotion mix. Business Horizons, 52(4), pp.357-365.

Hainla, L., 2018. 21 Social Media Marketing Statistics You Need to Know in 2018. Dreamgrow. Available at: https://www.dreamgrow.com/21-social-media-marketing-statistics/ [Accessed 4 May 2024].

KPM, 2013. Analiza e tregut të reklamës në Kosovë. Komisioni i Pavarur për Media – Republika e Kosovës. Available at: http://www.kpmks.org/materiale/dokument/1389360188.0656.pdf [Accessed 4 May 2024].

Kempt, S., 2018. Digital in 2018. We are social. Available at: https://wearesocial.com/blog/2018/01/global-digital-report-2018 [Accessed 4 May 2024].

Devika, M.D., Sunitha, C. & Amal, G., 2016. Sentiment Analysis: A Comparative Study On Different Approaches. Procedia Computer Science, 87, pp.44-49.

Heimerl, F., Lohmann, S., Lange, S. & Ertl, T., 2014. Word cloud explorer: Text analytics based on word clouds. In: 2014 47th Hawaii International Conference on System Sciences. pp. 1833-1842.

Hindle, A., Campbell, J.C., Hindle, A. & Stroulia, E., 2014. Latent Dirichlet Allocation: Extracting Topic. In: The Art and Science of Analyzing Software Data. Advances in Social Science, Education and Humanities Research, vol. 474, pp. 139-159.

Blei, D.M.(2012),"Surveying a suite of algorithmsthat offer a solution to managing large documentarchives", Communications of the ACM, Vol. 55No.4,pp.77-84.

Reisenbichler, M.and Reutterer,T.(2019),"Topicmodeling in marketing: recent advances andresearch opportunities",journal of BusinessEconomy, Vol.89 No.3, pp.327-356.

Chambers C, Feng Wc, Sahu S, Saha D (2005) Measurement-based characterization of a collection of on-line games. In: Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement, USENIX Association, pp 1–1

Yang, H.-E., Wu, C.-C. and Wang, K.-C., 2009. An empirical analysis of online game service satisfaction and loyalty. Expert Systems with Applications, [e-journal] 36(3), pp.1816-1825.

Holsapple, C., Hsiao, S.-H., & Pakath, R. (2014). Business Social Media Analytics: Definition, Benefits, and Challenges. Twentieth Americas Conference on Information Systems, Savannah.

Pasternak, O., Veloutsou, C. and Morgan-Thomas, A. (2017) 'Self-presentation, privacy and electronic word-of-mouth in social media', Journal of Product & Brand Management, 26(4), pp. 415-428. Available at: https://doi.org/10.1108/JPBM-04-2016-1150 [Accessed 4 May 2024]

eMarketer Inc. (2017) 'eMarketer Updates Worldwide Social Network User Figures', 17 July. Available at: https://www.emarketer.com/Article/eMarketer-Updates-Worldwide-Social-Network-User-Figures/1016178 [Accessed 4 May 2024].

Grimes, M. (2012) 'Global Consumers' Trust in 'Earned' Advertising Grows in Importance', 4 October. Available at: http://www.nielsen.com/us/en/press-room/2012/nielsen-global-consumers-trust-in-earned-advertising-grows.html [Accessed 4 May 2024].

Pozzia, F.A., Fersini, E., Messina, E. & Liu, B., 2017. Challenges of sentiment analysis in social networks: an overview. In Sentiment Analysis in Social Networks. Elsevier, pp. 1-10. Available at: http://dx.doi.org/10.1016/B978-0-12-804412-4.00001-2 [Accessed 4 May 2024].

Appendix

Link to Github with code

Dataset& Analysis: https://github.com/Serena1818/social-media-

: