The Research on the Social Media Analytica Model Apply on Barbie Culture Influences in Society

Haoxi Xie[1,a,*]

[1]*Faculty of Communication, University of Putra Malaysia, Selangor, Malaysia*
a. 208771@student.upm.edu.my
*corresponding author

Abstract: Barbie has long been one of the world's most influential icons for women; it was launched in 1959 by Ruth Handler. As a successful fashion doll model, she is a figurehead of the brand, leading females to have independence in life. Her influence has changed the social status and life of female groups. This paper starts with the data analysis of the first Barbie live-action movie in 2023. Then, it analyzes and predicts the number of plays on the Barbie YouTube channel. Use Python data analysis implementation to analyze text movie reviews. This Python data technique includes text analysis, visualization, sentiment analysis, topic modeling, keyword extraction, and machine learning prediction. Further, the author combine the analysis results and study the social background of the female group in different regions and discuss the influence of Barbie culture and women's social status in detail. Our side will further analyze the final result and obtained it's on society, obtain its impact on society.

Keywords: Barbie, female, data analysis, python

1. Introduction

Since its launch, Barbie has quickly become one of the world's best-selling toys and has made a huge cultural impact across the globe. Barbie is undoubtedly an iconic product in the history of toys, and its influence and popularity are unmatched globally. *Barbie* 2023 is a much more personal movie. Imbued with a sense of the film freedom and unfettered creative zeal throughout, the film contains so much rich and delicate detail - costumes and sets, gestures, and diction - that it is impossible to enumerate the many inventions and decisions that brought it to life. With its frenetic pace and expansive scale, it's like a live-action cartoon that captures the anything-goes spirit of classic Looney Tunes better than any other movie all have seen. The main idea is that Barbie comes back to life and enters the real world, but Gerwig makes this transition happen by cleverly giving Barbie a past life as a doll.

The Barbie doll played by Robbie is known as "Stereotypical Barbie" and lives in Barbie land with all the other Barbie dolls on the market, whether it's Astronaut Barbie, Dr. Barbie, or President Barbie, as well as Barbies of all races and sizes all types of people named Barbie, all living in doll houses [1]. Yet its whimsical plot is constructed with dramatic logic, successfully transforming fantastical leaps into persuasive results, with the result that the story's details seem completely inseparable from, and continuous with, the gorgeously decorative visual realm it sets in motion [2]. They can provide a strong sense of style on the screen, they are rich and overflow, because they should not be confined to the screen, but burst out to fill the whole space. The theater has taken its place all over the world.

She is not cynically borrowing from pop culture; The film simply borrows from popular culture, enthusiastically embraces it, and directly changes it. It changes the relationship between the audience and it. The relationship becomes active, critical, and non-nostalgic. The *Barbie* movie 2023 reinterprets the looming, sparkling cultural objects and artistic colors in society. Dramatize the link between playing in the children's playroom and, for progress, playing on the world's big screen.

Therefore, she is influential according to the above *Barbie* live-action movies and Barbie as the fashion leader of female groups in the image world [2]. Thus, this article will explain how to make detailed data analysis by Python through the box office income and regional data of *Barbie* live-action movies and generate the corresponding detailed data visual distribution map and heat map. It can provide researchers with more intuitive and precise objective results for analysis, and further analysis based on the income and cultural background of different box office regions and discuss the corresponding results and reasons to understand further the correlation between female culture and box office around the world through other box office incomes. Furthermore, through statistics, YouTube Barbie channel analysis and linear regression are built to predict the manifestation and direction of Barbie's influence more intuitively.

As a pretty pink phenomenon of Barbie's first live-on movie, its opening weekend they ranked 337 million US dollars, and another film helmed record made by women. With the guidance and exploration of director Greta Gerwig, the legacy of the iconic doll is used flexibly. Allowing Barbie to be manufactured as a feminist icon has unfortunately been criticized as a passive sexual target. The *Barbie* movie embodies Mattel's attempt to get a Barbie, a fashion icon, to play with her.

Then, the language and emotion analysis of Barbie's movie reviews are modeled to study the audience's emotions and thinking further. These methods can help author to have a more three-dimensional and intuitive insight into the standard data analysis and the influence presented by the current media and make it easier for writer to find and analyze some details that public have not noticed, which is the uniqueness of this article.

2. Research Method

2.1. Heat Map of Box Office Data of Barbie 2023 Movie

In 2023, the *Barbie* live-action movie, released worldwide, attracted many audiences to go to the theater to watch it. Through detailed analysis of the box office data collected by ourselves on the IMDb website, people can have an in-depth understanding of this movie's popularity and box office data in different regions and explore the reasons behind the following (Figure 1) [3].

2.2. Data Analysis of Barbie Channel Views

First, the "time" column is converted to a number, using it as an independent variable, and each point in time is simply assigned a consecutive integer. Then create a linear regression model using scikit-learn's Linear Regression class. The model is trained using the time number as input (X) and the number of views as output (y). The regression coefficient and intercept are obtained from the model. The regression coefficient (slope) is 1000, and the intercept is 12000. This means that for each additional unit of time, the number of views is expected to increase by approximately 1000. When the time is 0, the predicted number of views is 12000 (Figure 3).

2.3. Regression Equation Calculation Method

Sklearn's Linear Regression class is used to calculate the regression equation. The steps are as follows: Data preparation is when our group first convert the "time" column to a number so that ours can use it as an argument. Then the Linear Regression() function is used in the model creation to create a

linear regression model. Model training is to train the model using the fit() method, where the input is a time number and the output is the number of views. The coef_ and intercept_ attributes are used in the acquisition coefficient to obtain the slope and intercept from the model [4].

2.4. Machine Learning Model in Mean Square Error (MSE)

2.4.1. Data Segmentation

Our dataset first should divide the data into 2 set which are training set and test set. This way, ours can use the training set to train the model. The test set is then used to evaluate the performance of the model, so, choose 80% of the data for the training set and 20% for the test set [5].

2.4.2. Model Training

Ours trained the linear regression model using the data from the training set.

2.4.3. Evaluate the Model

To evaluate the performance of model accuracy ourselves choose to use the model to predict the number of views in the test set and calculate the mean square error (MSE) [6]. MSE is a common regression model evaluation index, which measures the mean squared difference between the model's predicted value and the true value. In this case, MSE is $4.15e+12$, which is a relatively large value.

2.5. Sentiment Analysis

Using the TF-IDF method, the following ten keywords were identified: 2023warning: 82.55, movie: 82.50, barbie: 68.36, film: 46.09, fun: 30.81, just: 29.35, good: 28.18, like: 27.64, great: 26.50, world: 26.01. The values of these keywords indicate their importance in the overall data set.

Keyword extraction analysis identified several keywords, such as 2023 warning, movie, barbie, and film. These keywords indicate that users refer to these topics frequently in their comments, and viewers care deeply about the clues.

Review length can give clues about user engagement with Barbie's movie. A longer review may mean more opinions or suggestions for the *Barbie* movie. The data showed that the length of the comments varied between 126 and 9,986 characters, suggesting that viewers of the 2023 *Barbie* movie had different levels of engagement when providing feedback (Figure 7). The length of most reviews is concentrated in a shorter range, which may mean that most moviegoers tend to provide short feedback [7].
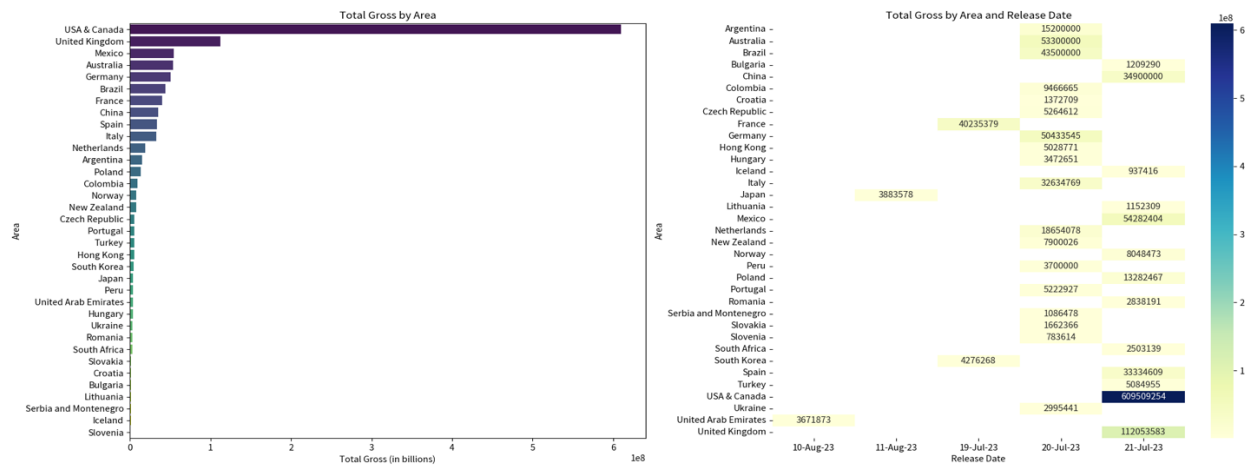
3.   Research Results

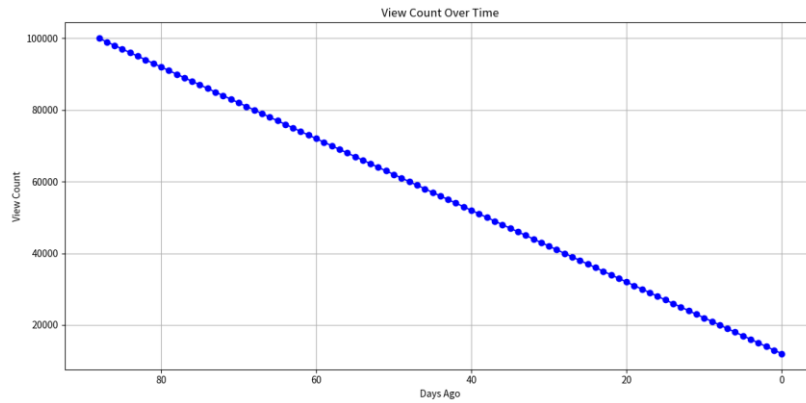Figure 1: Heat map of box office data of Barbie 2023 movie.

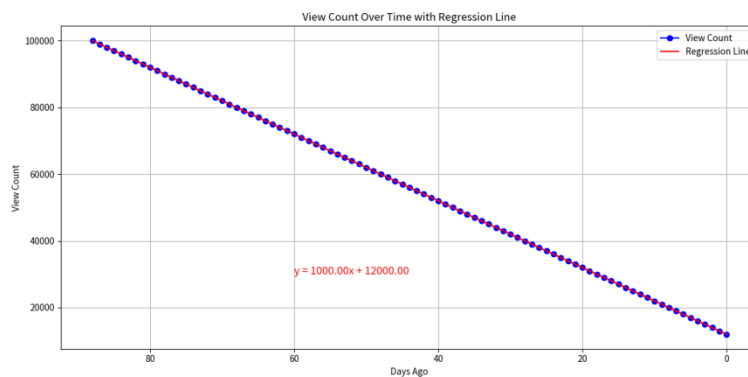Figure 2: Barbie youtube channel views.
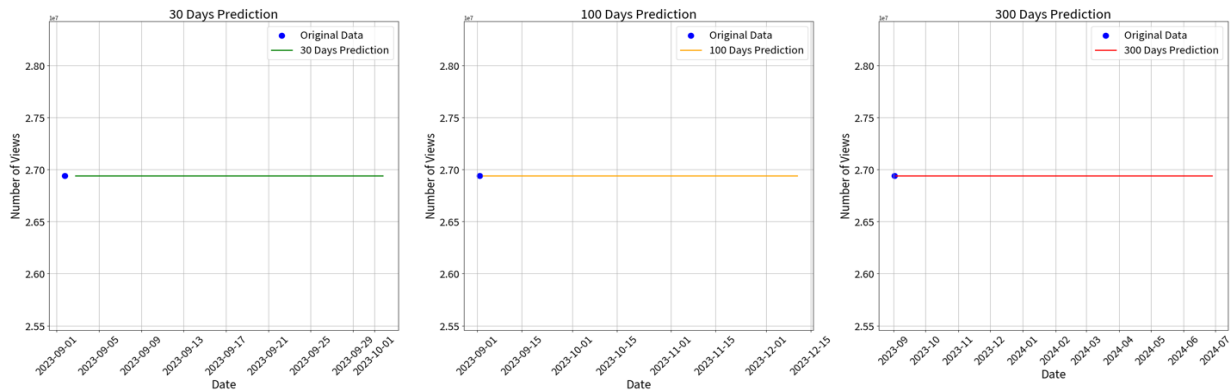

Figure 3: Regression equation.

Figure 4: Future prediction.

Use the model to predict future views. This can help ourselves understand future trends and possible changes.
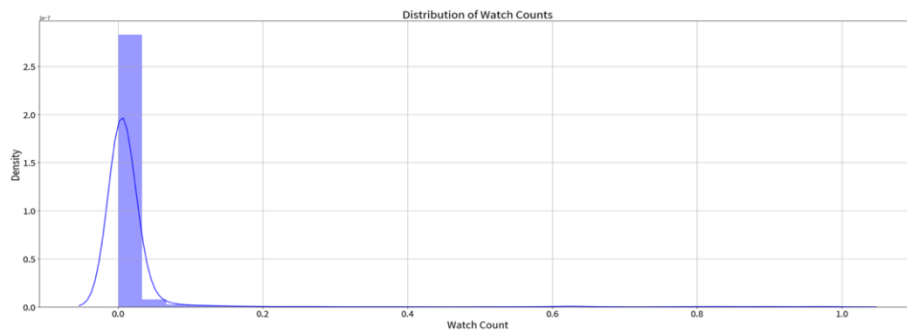

Figure 5: Data visualization.

The distribution of views shows a right-sloping pattern, indicating that most videos are viewed on the lower side, and a few videos have unusually high views. The distribution peaked at around 1 million, indicating that many videos were viewed close to that value. There are some videos that have more than 5 million views, which can be considered outliers or particularly popular videos (Figure 5).
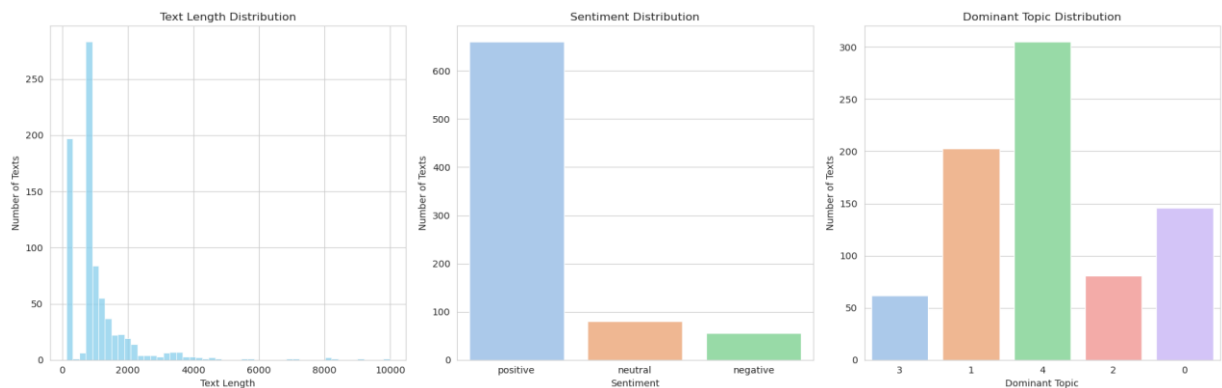

Figure 6: Text analysis, sentiment analysis, topic modeling, keyword extraction.
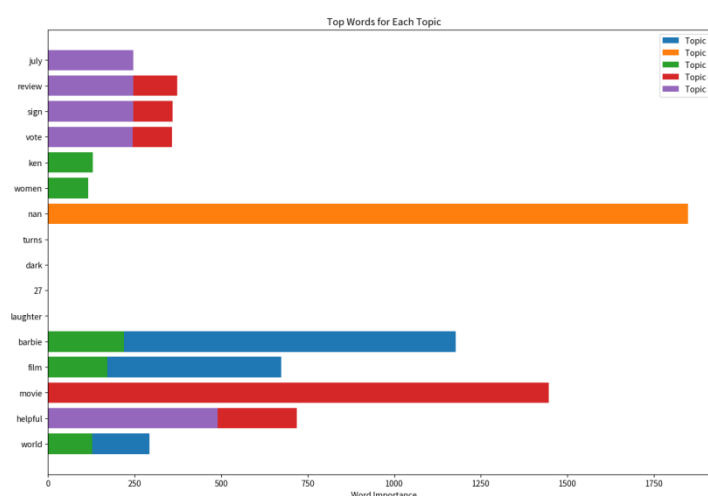
Figure 7: Sentiment analysis.

Text analysis, our analyzed the length of the comment text. The following are descriptive statistics for comment length Total comments: 797, minimum comment length: 126 characters, standard deviation: 1074.49 characters, 75th percentile: 1186 characters.

Sentiment analysis results: Positive comments: 660, neutral comments: 81, negative comments: 56, using Latent Dirichlet Allocation (LDA) method. Theme modeling also uses LDA method to identify the following five themes: Theme 1: audience, good, just, like, world, men, women, movie, barbie, film. Theme 2: perfect, ryan, real, gosling, margot, robbie, ken, world, movie, barbie. Topic 3: just, expected, spoilers3, spoilers4, barbie, men, women, great, film, 2023warning. Theme 4: think, funny, great, barbie, really, fun, good, like, just, movie. Theme 5: ryan, character, margot, design, gosling, robbie, greta, gerwig, film, barbie.

4. Research Discussion

4.1. Analysis to Box Office Data of *Barbie* 2023 Movie

First, it can be seen from the box office data that the box office of Barbie's 2023 live-action movie in the United States and Canada is far higher than in other regions, which is unsurprising because Barbie's origin country is the United States. For the United States and Canadian audiences, Barbie is a doll and a cultural symbol. The North American film market is relatively more mature, and the film promotion and distribution technology are outstanding, which is also an essential reason for the high box office (Figure 2), but it also reflects a paradox: in the place where the feminist movement originated and was most active, Barbie, who has been criticized for objectifying women, can still achieve such a high box office [8].

In the Middle East and South Asia, *Barbie* movies could have done better at the box office. Women's rights are limited in these places, and the image of Western women as independent and free, which Barbie represents, conflicts with local culture and values. Paradoxically, Barbie as a doll sold relatively well in those places. This mean local women yearn for more freedom and rights in private but are afraid to express them in public.

In addition, when the author further analyze the box office data of other regions, people will find some interesting phenomena. For example, in some countries with deep traditional culture, such as the Middle East and South Asia, the box office of *Barbie* movies could be better. This may be related to the status and rights of women in these regions. In some countries, women's rights and quality are

often restricted. As a symbol of female independence and freedom, Barbie may not be widely accepted by local audiences.

On the contrary, *Barbie* movies do relatively well at the box office in some countries with equal rights for women, such as some European countries. Because women in these countries tend to have an excellent educational background, they have a profound understanding of gender equality [9].

In general, the box office of *Barbie* live-action movies provides a window to observe the status and rights of global women. Through the analysis of these data, people can better and more deeply understand the situation of women in different countries and regions and their pursuit of gender equality [10]. As a symbol of global culture, Barbie undoubtedly provides valuable research material. While our study provides a snapshot of this impact in the context of a 2023 live-action film, the larger story of Barbie and its social impact is an ongoing narrative that sets the stage for future research and discussion(Figure 6). Dolls, in their myriad forms and interpretations, will no doubt continue to spark debate, inspire dreams, and reflect the ever-evolving tapestry of global society.

4.2. Data Analysis of Barbie Channel Views

First, a preliminary observation of the data was made. As can be seen from the chart, the number of views has shown an upward trend over time. This could mean that Barbie videos are becoming more popular over time, or it could be understood that the platform's user base is growing. While linear regression provides a simple prediction, it also highlights the inherent challenges of predicting digital content consumption patterns.

4.3. Regression Equation

$$\text{Number of views} = 1{,}000 \times \text{ time number} + 12{,}000 \tag{1}$$

The blue dots represent the actual number of views, while the red lines represent the predictions of the linear regression model. As people can see from the graph, the red regression line roughly matches the actual data points in blue, which means that the linear model fits the data well [11].

4.3.1. Observation Trend

It is clear from the graph that the number of views has shown an upward trend over time. This could mean that the content is becoming more popular over time, or that the platform's user base is growing.

4.4. Future Prediction

Models were used to predict the number of views over the next 30 days, 100 days, and a year.

Plot these predictions, along with the raw data and regression lines, so that you can visually see the effects of your predictions.

As shown in the figure above, the blue dots represent actual views, the red lines represent the predictions of the linear regression model, and the green "x" marks represent projected views over the next 30 days, 100 days, and one year (Figure 4).

4.5. Analysis

4.5.1. Regression Line and Actual Data

The red regression line roughly matches the actual data points in blue, which means that ours linear model fits the data well. However, due to the large mean square error, this also means that the model may not fully capture all the patterns in the data.

4.6. Future Predictions

The green "x" mark indicates future forecast views. As can be seen from the graph, the predicted views roughly match the regression line, which means that the future views predicted by the model are in line with the current trend [12].

4.7. Limitations of Models

While linear regression gives people an easy way to describe and predict data, it also has its limitations. For example, it assumes that the relationship between dependent and independent variables is linear, which may not be true in some cases. In addition, the model may be affected by outliers, resulting in inaccurate predictions (Figure 6) [12].

4.8. Function of Sentiment Analysis

Sentiment analysis can help researcher understand the emotional attitude of moviegoers. In the dataset, the results of sentiment analysis showed that many reviews were positive, which means that watching this movie was generally satisfied with Barbie. people can see that most users have a positive attitude towards Barbie-related products or services [13]. Of the 660 comments, most were positive, meaning users generally had a positive attitude toward Barbie-related content. This certainly provides a good market environment for a live-action *Barbie* movie in 2023. However, there were also 56 negative comments. These reviews may relate to certain inadequacies in the product, or certain users may have reservations about certain changes to the Barbie brand. For film producers, these suggestions and criticisms are valuable and can help them do better in future productions [14]. Sentiment analysis provides our term with a macro perspective to understand the public's emotion towards the movie and helps our group better understand the overall satisfaction of users and possible problems.

The results of the theme modeling also provided people with some interesting findings. Most of the comments focused on a few themes that could relate to aspects of Barbie's image, storyline, and interaction with the audience. This gives filmmakers a direction as to what audiences are most concerned about and what they should focus on when making a film.

Despite the relative equality of rights for women in Europe, *Barbie* films have failed to meet expectations at the box office. This may be because European women have realized that true female independence and freedom should not just stop at appearance and material things. They seek true equality, not the apparent independence that Barbie represents [15]. However, these negative comments are a crucial reminder that Barbie's image, while powerful, is not universally recognized or viewed uniformly. Themes and keywords drawn from the reviews emphasize the multifaceted nature of the film, from the storyline to the characters.

The ideas of female independence and freedom conveyed by *Barbie* films often coincide with the values of female audiences in these countries. In addition, ours can also see from the box office data that *Barbie* movies in some Asian countries, such as China, Japan, and South Korea, have also achieved good results. Women in these countries have made significant progress in the past decades, reaching a certain status in society, economy, and culture [15]. The ideas conveyed by the films align

with the modern women they aspire to, but at the same time, they also face enormous social pressure. The *Barbie* movie offers them a chance to escape reality but also reflects a contradiction: Women in these countries gain their rights and status in pursuing modernization and Westernization.

However, the contrasting box office performances in regions such as the Middle East and South Asia highlight the complex interplay between global cultural ICONS and local traditions and social norms.

5.   Conclusion

Barbie, whose origins date back to 1959, has been a cultural cornerstone and has had a significant impact on women around the globe, from a mere toy to a symbol of the feminine ideal. Barbie has experienced many aspects of social perception, criticism, and appropriation. In this study, the 2023 live-action film revitalizes Barbie's presence in popular culture and provides a lens through which to examine its evolving meaning in contemporary society. Through our Python data analysis, ours infer that the *Barbie* live-action movie is not just a movie exercise but reflects the dynamic relationship between culture and dolls. North America's massive box office success is another testament to Barbie's enduring appeal and cultural resonance in its home country. In addition, our analysis of the Barbie YouTube channel shows an upward trend in views, which indicates a growing interest in Barbie or an expanding user base.

On the other hand, the sentiment analysis of film reviews paints a more qualitative picture. Positive reviews have been overwhelming, and it's clear that the film has struck a chord with many. At its core, however, the success of the 2023 *Barbie* live-action movie and the ensuing discussion it sparked underscored a more extensive social narrative. Barbie has remained relevant despite its commercial origins and criticism of potential objectification. Whether seen as a symbol of female independence or criticized as a proxy for material success, Barbie reflects society's evolving views of women, ambition, and identity. As a cultural entity, people must recognize that Barbie has transcended its plastic limitations or cinematic image. Its impact on women's identity, aspirations, and perceptions of society is profound, multifaceted, and evolving.

This study has some deviations in data analysis and collection, which can not represent the whole, and the results of mechanical learning are less accurate because of the limited experimental samples. With the continuous updating of research technology, the accuracy of research analysis results will continue to increase. The constant updating and upgrading of research methods in future science will make it easier to analyze and gain insight into the different details of the past and obtain more incredible research results.

References

[1]   Robin,D. (2016)The Evolution of an Icon: A Comparison of the Values and Stereotypes Reflected in the Original 1959 Barbie Doll and the Curvy 2016 Barbie Doll.

[2]   Ahmad, H. (2020)Influencer of Barbie Doll Movies on Women in Pakistan.Global Multimedia Review, 3.1, 52-64.

[3]   Reilly,N.(2019).Women, Gender, and International Human Rights: Overview. International Human Rights of Women. Springer Nature Singapore Pte Ltd. 1-18. https://doi.org/10.1007/978-981-10-8905-3.

[4]   Aakansha,G.,Katarya,R.(2020) Social media based surveillance systems for healthcare using machine learning: A systematic review. Journal of Biomedical Informatics, Volume 108, https://doi.org/10.1016/j.jbi.2020.103500.

[5]   Boeun,K.,Ryu,K.H.,Heo,S.M.(2022)Mean squared error criterion for model-based design of experiments with subset selection.Computers & Chemical Engineering, Volume 159, https://doi.org/10.1016/j.compchemeng.2022.107667.

[6]   Bzdok,D.,Altman,N.,Krzywinski,M. (2018)Statistics versus machine learning. Nat Methods, 15(4),233-234. doi: 10.1038/nmeth.4642. Epub 2018 Apr 3. PMID: 30100822; PMCID: PMC6082636.

[7]   Zachlod, Cécile & Samuel, Olga & Ochsner, Andrea & Werthmüller, Sarah. (2022)Analytics of social media data – State of characteristics and application. Journal of Business Research, 144, 1064-1076. 10.1016/j.jbusres.2022.02.016.

[8]   Worobey,J.,Worobey,HS. (2014)Body-size stigmatization by preschool girls: in a doll's world, it is good to be "Barbie". Body Image, 11(2),171-4. doi: 10.1016/j.bodyim.2013.12.001. Epub 2014 Jan 4. PMID: 24394637.

[9]   Vairetti,C.,Eugenio Martínez-Cámara, Sebastián Maldonado, Victoria Luzón, Francisco Herrera.(2020) Enhancing the classification of social media opinions by optimizing the structural information. Future Generation Computer Systems, Volume 102,838-846. https://doi.org/10.1016/j.future.2019.09.023.

[10] Gehlenborg,N., Wong, B.(2012) Heat maps. Nat Methods 9, 213. https://doi.org/10.1038/nmeth.1902

[11] Jordan,MI.,Mitchell,TM.(2015) Machine learning: Trends, perspectives, and prospects. Science,349(6245),255-60. doi: 10.1126/science.aaa8415. PMID: 26185243.

[12] M.A.Al-Garadi et al.(2019) Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges. In IEEE Access, vol. 7, 70701-70718. doi: 10.1109/ACCESS.2019.2918354.

[13] Wnkhade, Mayur & Rao, Annavarapu & Kulkarni, Chaitanya. (2022) A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55. 1-50. 10.1007/s10462-022-10144-1.

[14] Xu, Q.W.,& Chang, Victor & Jayne, Chrisina. (2022) A systematic review of social media-based sentiment analysis: Emerging trends and challenges. Decision Analytics Journal, 3. 100073. 10.1016/j.dajour.2022.100073.

[15] Michaelidou, Nina & Micevski, Milena. (2018) Consumers' ethical perceptions of social media analytics practices: Risks, benefits and potential outcomes. Journal of Business Research, 104. 10.1016/j.jbusres.2018.12.008.