Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANALYSING HUMOUR AND SENTIMENT IN FAMILY GUY DIALOGUE USING NLP AND MACHINE LEARNING

PRATHIMA TIRUMALAREDDY[1], ESWAR REDDY KETHIREDDY[2], SIRISHA ALAMANDA[1], LAKSHMI SREENIVASA REDDY DIRISINAPU[3]

[1]Assistant Professor, Dept. of Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad, India

[2]Student, Dept. of Artificial Intelligence and Data Science, Chaitanya Bharathi Institute of Technology

[3]Associate Professor, Dept of Artificial Intelligence & Data Science, Chaitanya Bharathi Institute of Technology

E-mail: asirishanaidu@gmail.com

ABSTRACT

Trends and sentiments in the dialogue of Family Guy, a famous animated television show were analyzed. Using the Linear Discriminant Analysis (LDA) machine learning algorithm, prevalent topics and sentiments in the dialogue are identified. The relationship between humour in the dialogue and season-wise ratings is examined to provide insight into what makes Family Guy appealing to its audience. Results suggest that the show's humour, use of pop culture references, and social commentary contribute to its success. Additionally, sentiment analysis reveals that the show's humour is often characterized by sarcasm and irony. Four lexicons are used to identify positive and negative sentiment in the Family Guy dialogue. Three of the lexicons, Affin, Bing, and NRC, are existing, while the fourth lexicon is curated to analyse humour. The sentiment of the dialogue in each dataset is analysed, and results show that the sentiment varies depending on the characters and situations involved. The show often uses humour to address serious social issues. A comprehensive analysis of the dialogue in Family Guy and its implications for media and society is provided.

Keywords: *Sentiment Analysis, Dialogue, Lexicon, Humour, Linear Discriminant Analysis-LDA, Family Guy Dataset*

1. INTRODUCTION

In brief, this original research paper aims at establishing a relationship between humor, emotions (and other sentiments and feelings) as a function of season and time. Insights into the reason and facts that made the animated television shown, namely, the family guy, famous, were provided. To this end the machine learning algorithm, namely, the linear discriminant analysis (LDA) is used for analysing the trends and sentiments in the dialogue of the "family guy".

A web of science search with the keywords, namely, linear discriminant analysis yields 20, 288 results as on 1st March 2025. The results shrink down to by two orders of magnitude to 232 when the keywords, namely, linear discriminant analysis and media studies were crossed. The results further shrank down to only 3 when the keywords, namely, linear discriminant analysis and media studies and sentiment analysis were crossed [1-3]. This shows the originality of the current work that is related to the use of linear discriminant analysis for sentiment analysis in media studies.

Television shows and movies have the power to reflect and influence society's attitudes and emotions. Family Guy, an animated television show created by Seth MacFarlane, has been entertaining audiences since 1999 with its unique brand of humour and social commentary. The show's success has made it a cultural phenomenon, and it continues to shape societal norms and influence popular culture.

This study analyses trends and sentiments in the dialogue of Family Guy to gain insight into the show's appeal to its audience. The LDA machine learning algorithm is used to identify prevalent topics and sentiments in the show's dialogue, with the help of the topic allocated, humours topics with highest probability are used to create humour lexicon and the relationship between humour in the

dialogue and season-wise ratings is examined to understand the factors that contribute to the show's success.

Along with this, impact of each character of the show is calculated by calculating their individual humour ratings on each season, which reveals most exciting characters of the show and dullest characters of the show.

Furthermore, a sentiment analysis is performed on two different datasets of Family Guy dialogue, which were created and pre-processed. Three existing lexicons are used to identify positive and negative sentiment, and the sentiment of the dialogue in each dataset is analysed. By doing so, the study explores how the show uses humour to address serious social issues and how the sentiment of the dialogue varies depending on the characters and situations involved.

This study contributes to the growing field of sentiment analysis and media studies, demonstrating the value of using data-driven approaches to understand popular media. By analysing the dialogue of Family Guy, insight is provided into the show's impact on popular culture and its appeal to viewers. Furthermore, the methodology and the lexicon curated for analysing sentiment in dialogue can be applied to other forms of media, making this study valuable for researchers in the fields of media studies, sentiment analysis, and computational linguistics.

2. LITERATURE SURVEY

In [4] the authors proposed a method for predicting humour response in dialog using acoustic and language features. The data used in the study is sourced from two popular TV sitcoms. A Conditional Random Field is used as the classifier/predictor due to the sequential nature of humour response in dialogues. The proposed method shows relative effectiveness with a maximum precision of 72.1% and 60.2% for two sitcoms respectively. The experiments conducted reveal that audio, speed, word, and sentence length features are the most effective. The main drawback of this is, it requires supervised dataset and, in most cases, the huge number of dialogues makes it complex to use these methods.

To address this drawback, the authors of [5] proposed a model for sarcasm detection based on utterances of various words. A comparative analysis is carried out against existing approaches, revealing that this method achieves better performance than the previous models with over one F1-score point improvement in sarcasm detection and ten F1-score points in humour classification. This analysis is done on comics which have old style of humour, so this method is needed to adapt for new kind of humour by creating a new lexicon. Creating new words for humour is also a hectic task as it requires going through the whole text corpus manually.

In order to create humour lexicon, we can make use of topic modelling algorithms and get a list of words which are probable for topics like irony, sarcastic etc. The authors of [6] proposed one such topic modelling algorithm, it helps to identify themes or topics within the text, which include themes like humour, irony etc. We can combine both topic modelling algorithms which helps to create lexicon and natural language processing techniques like sentiment analysis to identify humour.

The authors of [7] proposed a lexicon-based approach to sentiment extraction from text is presented, whereby a dictionary of words annotated with their polarity and strength, along with intensification and negation, is used. This approach is employed for the polarity classification task, which involves assigning a positive or negative label to a text based on its opinion towards the primary subject matter. The proposed method demonstrates consistent performance across domains and on unseen data. With reference to this paper lexicon-based humour analysis can be done similar to sentiment analysis.

The authors of [8] proposed a multimodal architecture for humour analysis using transformer-based architecture which achieved state of art performance. To train this model, a supervised dataset is required which is not possible in all the cases.

With the advent of Transformer based architectures, we may have a doubt about whether lexicon-based approaches work as effectively as transformers. The authors of [9] and [10] compared both the approaches and concluded that the transformer-based architectures slightly perform better than lexical-based approaches, but the problem here is the computational complexity of the transformer-based approaches. For huge dataset processing every dialogue can be hectic for slightly higher performance. They concluded that lexicon-based approaches can still be used with lower computational requirements.

The authors of [11] conducted a study on usage of lexical-based approaches and supervised learning approaches and concluded that lexical based approaches outperform the supervised learning approaches.

To understand various other sentiments, the authors of [12] proposed NRC lexicon which identifies several other emotions and authors of [13] and [14] proposed comparative studies on three lexicons, NRC, Bing, and Afinn. With the help of other sentiments nuances in humour can be understood like whether the humour is based in a serious scene, or some ironic scene etc.

The authors in [15] presented a study on humour in the form of puns, basically identification of pun location and translation on the JOKER dataset. The authors of [16] focused on multimodalities involving both languages English and Chinese where the dialogues are drawn from two sitcoms. In [17] the authors attempt to classify the humour into either offensive or non-offensive and reported statistically distinctive features representing both classes. In [18] the authors contributed a corpus of dataset which consists of explainable Chinese humour dataset. Their work also improved the humour generation response by pretrained models. In [19] the authors worked on tweets in Spanish language to automatically detect humour, their model's robustness is attributed to its ability in capturing the linguistic structure and sentence embeddings.

The research objectives of the work were to analyze the trends and sentiments in the dialogue of Family Guy, identify prevalent topics, sentiments, and the relationship between humour and season-wise ratings. This work also aims to explore how the show uses humour to address serious social issues, understand the sentiment of the dialogue based on characters and situations, and provide insight into the show's impact on popular culture and its appeal to viewers The study also sought to develop a custom humour lexicon and analyze the sentiment of Family Guy dialogues using existing lexicons such as NRC, Afinn, and Bing, which to the best of our knowledge is not covered in the literature studied by us.

3. METHODOLOGY AND RESULTS

3.1. Datasets

There are seven datasets used for the analysis of humour and sentiments of the dialogues in the family guy sitcom. The first dataset [20] used is sourced from Kaggle and contains the dialogues from the Family Guy spanning over 11 seasons. The dataset [20] has three attributes - character, dialogue, and season. The character attribute refers to the name of the character who spoke the dialogue, the dialogue attribute contains the text of the spoken dialogue, and the season attribute denotes the season number to which the dialogue belongs. This dataset [20] is in a CSV format and is readily available on the Kaggle platform.

```
220
00:10:03,122 --> 00:10:05,924
I'm the one who had to throw
away the damn time machine.

221
00:10:05,959 --> 00:10:07,326
Wait a minute. That's it!

222
00:10:07,360 --> 00:10:09,495
I've just got to rebuild
the time machine,

223
00:10:09,529 --> 00:10:11,663
and then I can go back
and save Brian!

224
00:10:14,901 --> 00:10:17,636
Oh, happy birthday, Neil!
```

Figure 1: Raw text from subtitles without any abrupt breaks

The humour lexicon is validated by comparing it with other three lexicons and showed positive results. The humour rating given by the lexicon is scaled to the GPT Humour ratings of each episode using linear regression [23] and visualized the correlation between context based deep learning models like GPT 3.5 and the lexical models.

The second dataset [24] used in the work is a custom dataset created by us. The subtitles for each episode are sourced from [21] and by making use of

a python script we extracted dialogues from the srt files. The script reads the subtitles from each episode, parses the data, and stores the relevant information, including the dialogue text, episode, timestamp, and season numbers, in a CSV file. The script is designed to automate the process of extracting the subtitles and create a comprehensive dataset that can be used for analysis. The subtitles contained various symbols and the dialogues abruptly ended without any continuity like shown in Figure 1 which made the dataset unusable initially and then it is pre-processed as shown in the Figure 2.

12	6	00:10:03	I'm the one who had to throw away the damn time machine.
12	6	00:10:05	Wait a minute. That's it!
12	6	00:10:09	I've just got to rebuild the time machine, and then I can go back and save Brian.
12	6	00:10:14	Oh, happy birthday, Neil!

Figure 2: Converted Text Without Any Abrupt Breaks

The process of converting raw subtitles to actual dataset contains several methods, firstly we need to remove extra symbols which indicate background scores, music etc. This is done by removing any Unicode characters present in the text. To ensure dialogue continuity, we identified patterns at which the dialogue is continuation from previous text. For example, in figure 1, the dialogues which are uttered at different time frames have small letters at starting. In few episodes it is indicated by symbols like hyphen (-) and few other Unicode characters.

The curated dataset contains the dialogues from 20 seasons spanning over 351 episodes of the TV show. The dataset has four attributes - season, episode, timestamp, and dialogue. The season attribute indicates the season number, the episode attribute indicates the episode number, the timestamp attribute denotes the timestamp of the dialogue in the episode, and the dialogue attribute contains the text of the spoken dialogue. The total number of dialogues in this dataset is 155,243.

The third dataset is the GPT 3.5 humour ratings of each episode of every season. The data is collected by prompting GPT 3.5 turbo 0125 [22] model to give ratings in the scale of 1 to 10. This data is useful for validating lexical analysis of humour against context-based humour recognition models. It contains 3 attributes. The first one indicates the season number and the second attribute contains the episode number. The third attribute contains the Humour ratings given by [22].

Another four datasets are used for the validation of the work which includes futurama scripts dataset [25], Southpark lines [23], Seinfeld Dialogue Dataset [27] and The Simpsons Dataset Transcript [28]. Each of the datasets contains attributes like dialogues, season and episode number and some other relevant information like the name of the episode, time stamp at which dialogues uttered etc. For all of these datasets, Humour ratings are assigned by GPT 3.5 turbo 0125 [22] model episode wise and the humour ratings are compared against the humour ratings predicted by the lexicons.

The research followed a structured methodology comprising data collection, preprocessing, sentiment analysis, lexicon development, and validation. The primary dataset, sourced from Kaggle, includes Family Guy dialogues across 11 seasons with attributes such as character, dialogue, and season. After cleaning and preprocessing to remove redundancies, stop words, and unwanted symbols, the data was prepared for analysis. Sentiment analysis was then conducted using multiple lexicons—NRC, Afinn, Bing, and a newly developed humour lexicon. The custom humour lexicon was created by annotating over 5,000 words derived from LDA-generated topics, with humour scores assigned using GPT-3.5 to capture contextual nuances. Finally, the model and lexicon were validated by applying them to dialogues from other sitcoms, ensuring the robustness and generalizability of the findings.

3.1.2 Preprocessing

There are several pre-processing tasks carried out to ensure the greater correlation between humour ratings predicted and context based GPT Humour ratings. These pre-processing tasks can be carried out for any text corpus in natural language processing for sentiment analysis.

One of the primary pre-processing tasks in natural language processing is stop word removal. Stop words are commonly used words in a language that do not carry any specific meaning, such as "the," "there," and "a." In the case of analysing the dialogue in Family Guy, stop words need to be removed to eliminate redundancy and improve the accuracy of the analysis. By removing stop words, the focus can be on the most meaningful words and phrases in the dialogue, which can help identify important topics and sentiments.

The next task is to remove any redundancies present in the text, the text may contain redundancies like numbers, extra spaces, and special symbols. These redundancies can also affect the

accuracy of the analysis, so they need to be eliminated as well. By removing these redundancies, the analysis can be focused on the most meaningful aspects of the dialogue.

Another pre-processing task that can improve the accuracy of the analysis is lemmatization. This task involves reducing words to their base form, or lemma, which can help eliminate redundancy caused by similar words like "jump," "jumps," "jumping," and "jumped." By lemmatizing words, the analysis can be focused on the most meaningful words and phrases, regardless of their tense or other grammatical variations as word used in any tense or form may represent same meaning.

For the second dataset, the dialogues analysed in Family Guy are extracted from subtitles, which means that a single dialogue may be split across multiple frames. This can make it difficult to analyse the dialogue. To address this issue, the multiple dialogues need to be merged to ensure that each dialogue is contained within a row. By doing so, each dialogue can be analysed in its entirety and a more accurate understanding of its meaning and sentiment can be obtained.

Finally, unwanted symbols, Unicode characters and notations like music notes and tags need to be removed. These symbols can interfere with the analysis and make it more difficult to identify important topics and sentiments in the dialogue. By removing these symbols, the analysis can be focused on the most meaningful aspects of the dialogue and the sentiment conveyed by each dialogue can be accurately identified

In the actual process for sentiment analysis, we make use of several lexicons to calculate the sentiment of each dialogue.

The methodology of this work is represented in figure 3, where the datasets are pre-processed and used for sentiment and humour analysis. Humour ratings are assigned by two different ways, using GPT and using Lexicon. The ratings given by GPT are taken as baseline and the lexicon ratings are compared against it to find whether lexical analysis still works in the era of Large Language models. The curated lexicon is validated by using it against other sitcoms similar to The Family Guy with the humour ratings given to them by the same GPT model

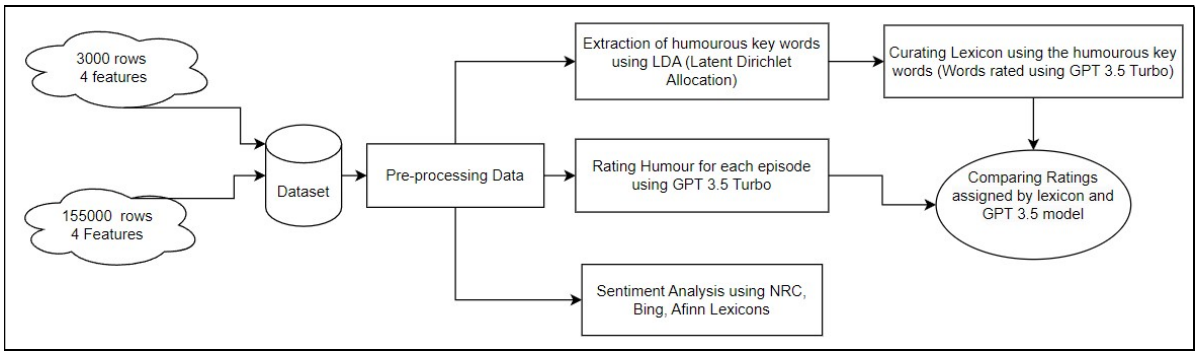

Figure 3: Methodology Of The Study

3.1.3 NRC lexicon

The NRC lexicon [9] is utilized to predict the sentiment of Family Guy dialogues. Emotions such as anger, depression, and excitement are associated with a list of words in this lexicon. These emotions are extracted for each dialogue and utilized as features to forecast the sentiment. The sentiment of each episode is analysed to comprehend the emotional content and themes.

According to Figure 4, with the words generated according to each topic from NRC lexicon, we can know that most of the humour generated is by sentiments like disgust, joy, surprise etc, as they have words like damn, fat, shot, fun etc. which are also predicted by LDA model as humorous topics.

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

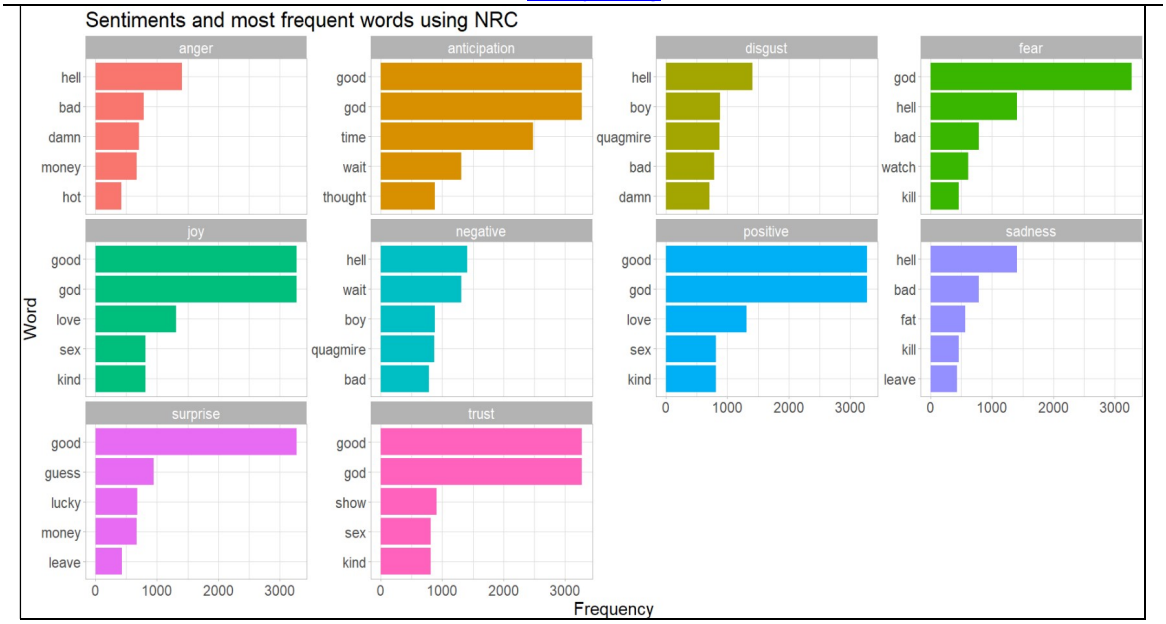

Figure 4: Most Repeated Words According To Each Sentiment Generated Using NRC Lexicon.

3.1.4 Afinn lexicon

The Afinn lexicon is employed to predict the sentiment of Family Guy dialogues. Each word in the dialogue is given a sentiment score ranging from -5 to 5 by the Afinn lexicon, which is used to calculate the overall sentiment of the dialogue. This technique is used to determine the overall tone of the dialogue, and the sentiment of each season is analysed to identify the themes and emotions.

3.1.5 Bing lexicon

The Bing lexicon is employed to categorize the sentiment of Family Guy dialogues into positive and negative classes. This lexicon consists of a list of words and phrases that are classified as positive or negative. The sentiment of each season is analysed to understand the overall sentiment and identify any patterns or trends.

In the Figure 5, the sentiments of each season are calculated using Bing and afinn lexicon. In Bing lexicon, the sentiments are directly labelled as positive, negative or neutral whereas for afinn lexicon the sentiments are rated from -5 to +5, so we labelled ratings greater than zero as positive and ratings less than zero as negative.

Bing lexicon shows 50408 dialogues as positive, 35760 dialogues as negative and rest others as neutral. Similarly, afinn lexicon shows 56895 dialogues as positive, 36385 dialogues as negative and rest others as neutral. The higher number of positive utterances suggests that Family Guy tends to maintain an overall positive tone throughout its seasons. This positivity could be attributed to the show's comedic nature, which often relies on humour and satire to entertain its audience.

While there are the greatest number of positive dialogues according to Figure 4, there are also significant number of negative dialogues. In the context of the show, negative utterances are due to involvement sarcasm or irony, where characters express dissatisfaction or critique in a humorous or satirical manner. These instances of humour could contribute to the overall negative sentiment while still being perceived as comedic by the audience

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

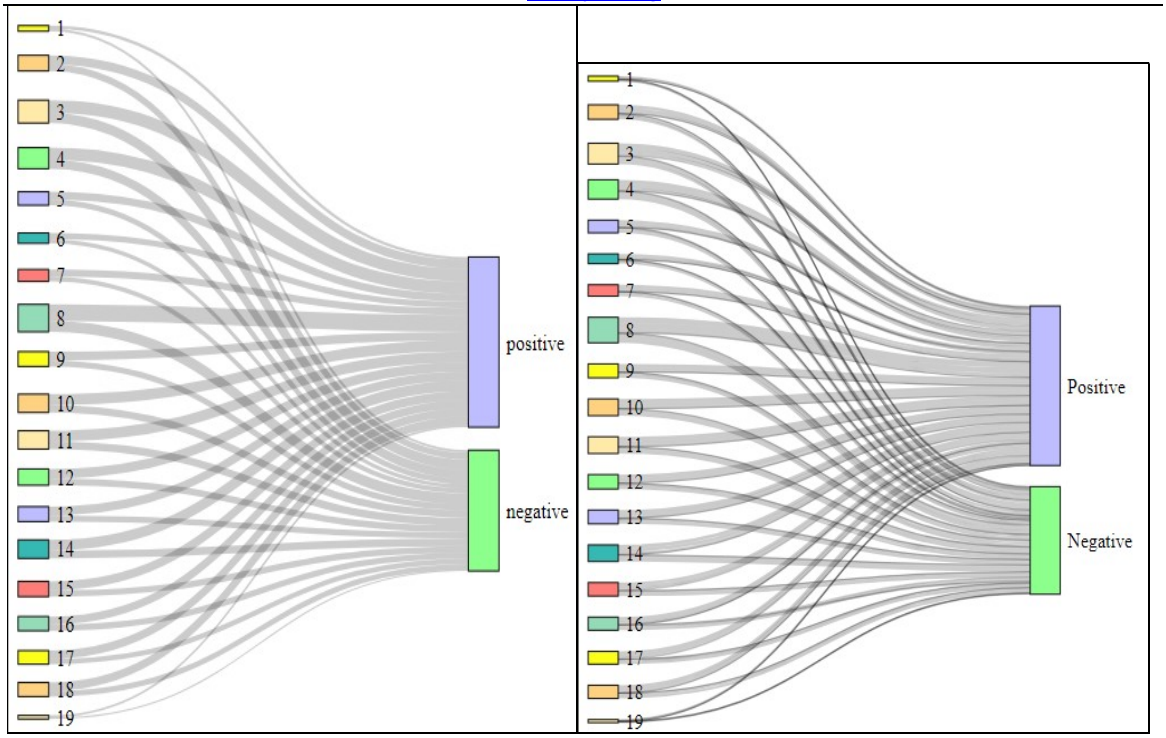

Figure 5: Sentiments Of Bing And Afinn Lexicons Calculated By Each Season.

In the Figure 5, the sentiments of each season are calculated using Bing and afinn lexicon. In Bing lexicon, the sentiments are directly labelled as positive, negative or neutral whereas for afinn lexicon the sentiments are rated from -5 to +5, so we labelled ratings greater than zero as positive and ratings less than zero as negative.

Bing lexicon shows 50408 dialogues as positive, 35760 dialogues as negative and rest others as neutral. Similarly, afinn lexicon shows 56895 dialogues as positive, 36385 dialogues as negative and rest others as neutral. The higher number of positive utterances suggests that Family Guy tends to maintain an overall positive tone throughout its seasons. This positivity could be attributed to the show's comedic nature, which often relies on humour and satire to entertain its audience.

While there are the greatest number of positive dialogues according to Figure 4, there are also significant number of negative dialogues. In the context of the show, negative utterances are due to involvement sarcasm or irony, where characters express dissatisfaction or critique in a humorous or satirical manner. These instances of humour could contribute to the overall negative sentiment while still being perceived as comedic by the audience. With this we can say that a significant amount of humour is generated from negative sentiments.

Using Latent Dirichlet Allocation for Humour analysis:

LDA (Latent Dirichlet Allocation) is a topic modelling algorithm which discovers various abstract topics within the collection of documents. While, it does not predict the humour directly, it will list the words for the topics like humour, irony, sarcasm and other irrelevant topics. Overall, the LDA model was used for exploratory data analysis to uncover the underlying themes in the dialogues of the TV show. The words of these topics can be used to create a lexicon. We created a lexicon of 5065 words to analyse the humour. For each word humour score is assigned in the scale of 1 to 10 using GPT 3.5 turbo 0125 as illustrated below. For example, in Figure 6, the word humiliate is given 1 because this word is mostly uttered in few serious conversations and generate no humour, and the word kidding is given a rating of 5 because the character 'Greg' will utter this word after he played a satire on other characters by saying the phrase 'just kidding'. The word awful is given highest possible rating because when some character uttered 'awful', they are feeling uncomfortable which generates visual kind of humour. This also explains why there are a significant number of negative dialogues in the show according to the sentiments classified by Bing and Afinn lexicon in the Figure 5. The words present in the lexicon are

mostly based on the pop culture references and satires on current affairs as they generate most amount of the humour for family guy sitcom.

humiliate	1	kidding	5	awful	9

Figure 6: Humour Scores Across Various Words From 1 To 9

This lexicon is utilized to calculate the humour score for each dialogue, and the humour content and themes across each episode are analysed. The average of humour for each episode is calculated using mean of the humour of each dialogue.

$$\text{Humour rating of an episode} = \frac{\text{sum of humour ratings of dialogues in an episode}}{\text{total number of dialogues in an episode}}$$

Using Linear Regression [20], the relationship between the predicted humour ratings and IMDB episode-wise ratings is examined to provide insight whether humour is making Family Guy appealing to its audience.

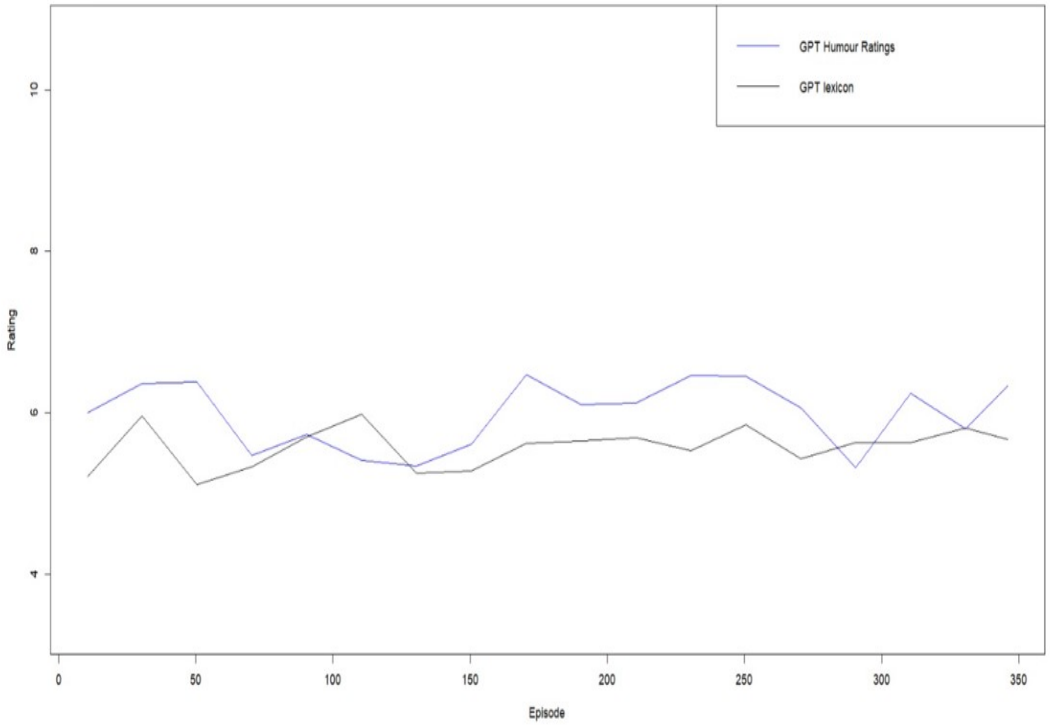

Figure 7: The Correlation Between Lexicon Humour Ratings And GPT Humour Ratings

The relationship between season-wise GPT humour ratings and humour content in Family Guy dialogues is analysed in figure 7. Humour scoring using lexicon is utilized to understand the correlation between GPT humour ratings and the show's humour content. The humour content of each episode is examined to understand how it impacts the overall quality and success of the show. The ratings for each episode are calculated and each point in the line plot is plotted for every 20 episodes to understand the correlation between humour and user ratings clearly.

If we observe Figure 7, there is a comparison between Humour ratings given by lexicon and the context-based humour ratings given by GPT 3.5 model [14]. If we observe the graph, both the ratings are moving in a similar way (ups and downs in the graph). The analysis using the lexicon can be performed with low number of computational resources whereas the resources consumed by [22] to perform context-based humour analysis are enormous. For instance, it took about an hour for GPT through OpenAI API to assign ratings for all the 351 episodes (around 1 million tokens) which is quite high compared to less than a minute time taken for lexical based analysis.

Similar kind of process is followed for the [25], [26], [27] and [28] datasets and compared the humour ratings given by the lexicons created from the respective text corpus and the humour ratings given by the family guy lexicon [24].

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

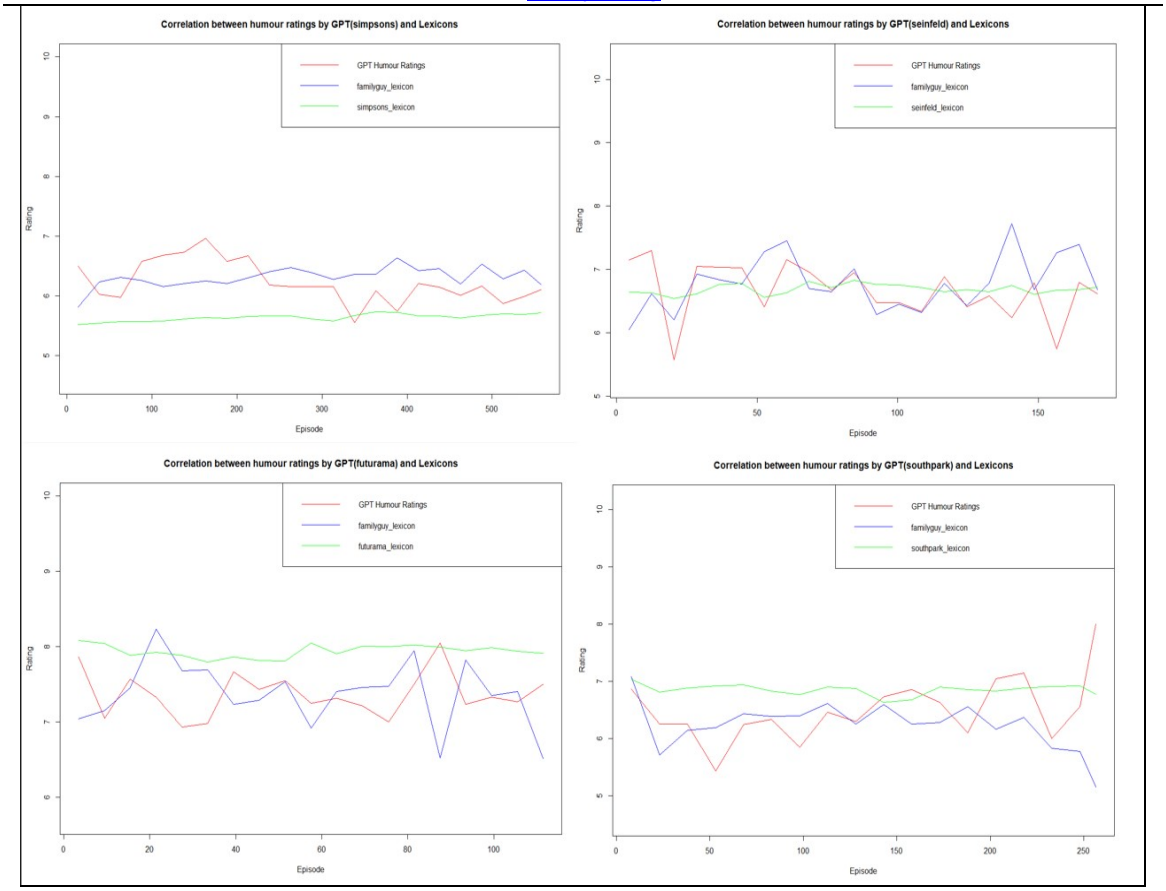

Figure 8: Comparison Of Humour Ratings Given By Respective Script's Lexicon And The Family Guy Lexicon With GPT Rating As Standard

Figure 8 demonstrates that there is greater correlation between GPT humour ratings and family guy lexicon humour ratings than GPT humour ratings and the respective lexicon humour. This is due to greater variety of words uttered in the family guy tv show and the humour generated through wordplay in the show. So, we can confidently say that the lexicon curated for the family guy will work for most of the similar kind of sitcoms without consuming much computational resources.

The table 1 below represents the mean absolute error calculated between the Ratings of GPT, respective lexicons and the family guy lexicon. The data proves that family guy lexicon is sufficient for the lexical based analysis of any sitcoms, in fact the mean absolute error by the family guy lexicon is less compared to humour ratings of their own lexicons. So, the same lexicon [24] can be used for any other similar kind of humorous shows.

Table 1: Mean Absolute Error Between GPT Ratings And Lexicon Ratings

Dataset	***Dataset's Lexicon***	***Family Guy Lexicon***
Family Guy [24]	0.53	0.53
Simpsons [25]	0.59	0.38
Seinfeld [26]	0.42	0.40
Futurama [27]	0.57	0.46
Southpark [28]	0.56	0.53

For Simpsons [25] and Futurama [27] datasets, there is significant difference between their own lexicon and the family guy lexicon. This is due to both shows are created by the Matt Groening for Fox Production Company. The same company which produced the Family Guy TV show and the creators of all three are in the same company and in fact they made crossovers between the shows which demonstrates that these shows are most similar to family guy, So the lexicon have greater impact on these shows.

Journal of Theoretical and Applied Information Technology
31st December 2025. Vol.103. No.24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Along with these, Figure 9 represents each character impact by season across 11 seasons of Family Guy based on the humour content of each season by each character. The relationship between the humour content and the involvement of different characters is examined to identify any trends or patterns. We can also understand the nuances of each character based upon their humour.

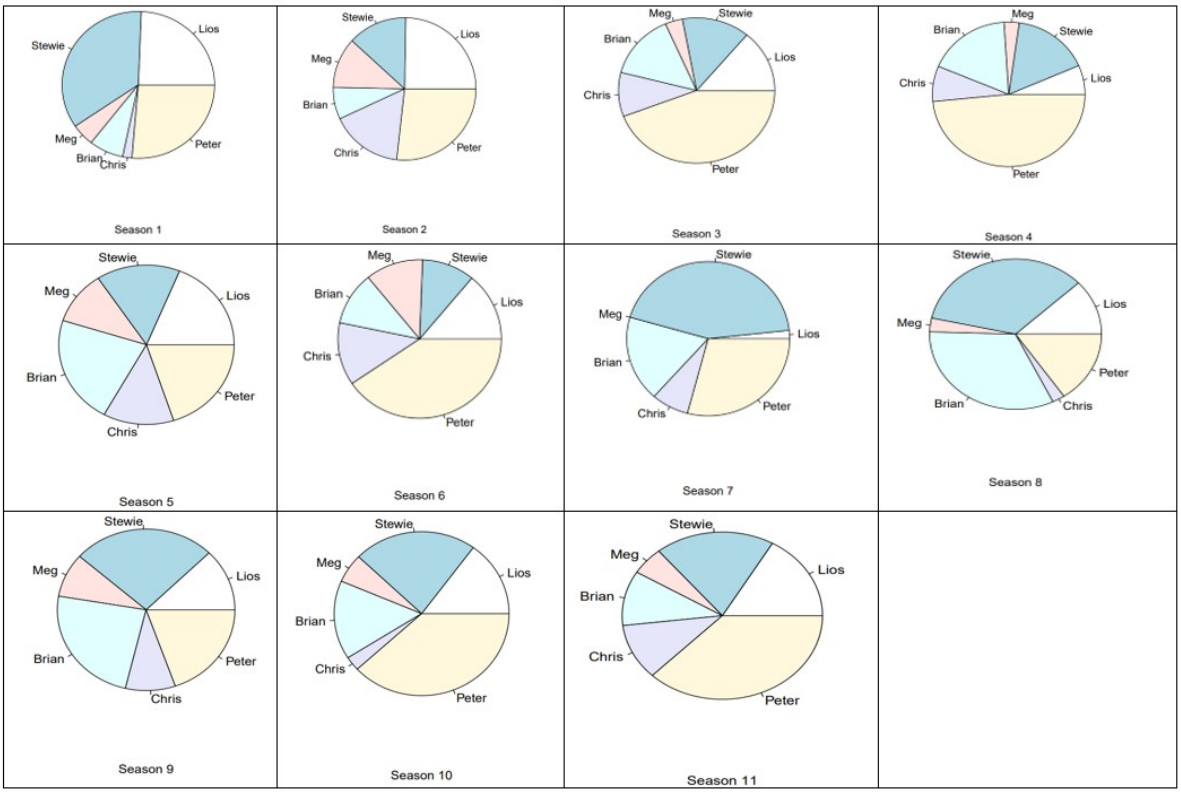

Figure 9: Impact Of Each Character On Humour Ratings By Each Episode

From the figure 9, we can understand that there is a dominance of single most character Peter Griffin, who is the main lead of family guy generated most amount of humour. This fact is accepted by everyone who watches the show.

The study advances the current state of the art by applying NLP and machine learning techniques to analyze humour and sentiment in *Family Guy* dialogues. It uniquely explores how humour, emotions, and sentiments evolve across seasons, revealing how character dynamics, social commentary, and contextual humour contribute to the show's popularity. Distinct from prior work, this research introduces a custom humour lexicon and employs multiple sentiment analysis approaches, including deep learning models, to capture nuanced, context-based humour. By bridging computational methods with media studies, it provides fresh insights into the cultural impact of *Family Guy* and establishes a foundation for future research in humour and sentiment analysis.

4. CONCLUSION

The study uncovers the nuances present in sentiments and humour present in Family Guy show with lexical based analysis. By using NRC, Bing, Afinn lexicons to understand the sentiments of each character and understand which kind of sentiments contribute to humour.

We proposed a new lexicon for humour analysis by annotating 5065 words with humour ratings ranging from 1 to 10. It is developed using the words taken from LDA algorithm and annotating the humour score for each word by GPT 3.5 [22] with enough context. The lexicon was developed to supplement the existing NRC, Afinn, and Bing lexicons, to understand the nuances between the sentiments of the dialogues and humour present in each dialogue. To assess the humour in the dialogue, the study

analysis using the custom humour lexicon and plotted the season-wise humour scores using a line graph. This analysis provided insight into the relative levels of humour in each season and allowed the identification of trends and changes in the use of humour throughout the show's run. Additionally, the study examined the relationship between lexical based humour ratings and GPT humour ratings, using the ratings generated by [22]. By comparing the humour scores to the GPT ratings, the study was able to identify that the context-based humour ratings and lexicon-based humour ratings are actually correlated, and gain insight into whether transformer-based architectures are performing for the given computational resources. This shows that the work done by [9] and [10] by comparing deep learning approaches to lexical based approaches still holds true for majority of the shows. Similar approach is followed for another 4 similar sitcoms and evaluated the performance of the family guy lexicon against their respective humour lexicons curated using same process. The performance of family guy lexicon is better than their respective lexicons when compared to GPT ratings given for the text corpus for every episode. Finally, the study analysed the impact of each character on each season, using a pie chart to display the proportion of dialogue spoken by each character in each season. This analysis allowed the identification of which characters had the most influence on the dialogue and humour in each season and provided a deeper understanding of the show's character dynamics. By examining the dialogue using multiple sentiment analysis lexicons, the authors provide a more nuanced understanding of the show's use of humour and its impact on popular culture. This methodology for sentiment analysis can be applied to other forms of media, making this study valuable for media studies and sentiment analysis.

In conclusion, the study successfully achieved its objectives by conducting sentiment analysis on two different datasets of Family Guy dialogue, exploring the show's use of humour to address serious social issues, and understanding the sentiment of the dialogue based on characters and situations involved. The study also developed a custom humour lexicon and analyzed the sentiment of Family Guy dialogues using existing lexicons, providing a more nuanced understanding of the show's use of humour and its impact on popular culture.

5. FUTURE SCOPE

There are several potential avenues for future research and development based on the findings of this work. Firstly, the custom humour lexicon could be expanded and refined to improve accuracy in detecting humour in dialogue. Additionally, this lexicon can be used in combination to Deep Learning based models like BERT to understand context-based humour and further increase the relevance of humour ratings to GPT ratings. Similar kind of lexical analysis can be done for other genres like action, drama, adventure and can be used for classification of various shows based on the genre they fall in. This process has many real time applications like the recommendation systems for OTTs which require lower computational complexity.

Further research could also explore the impact of humour on different audience segments to better understand the show's appeal to different demographics. Additionally, analysis of humour in other media formats could reveal insights into the use of humour across different contexts.

Based on the findings of this study, there are also practical applications in marketing and content creation. A humour analysis tool could be developed to provide insights for businesses and content creators looking to optimize the use of humour in their campaigns.

Finally, research into cross-cultural humour analysis could provide valuable insights into how humour is used and interpreted across different cultures and languages. Overall, the potential for further exploration and development in the field of humour analysis is vast, and this work serves as a foundation for future research and innovation. The study's focus on Family Guy dialogue may limit the generalizability of the findings to other media formats or cultural contexts. These limitations should be considered when interpreting the results and applying the study's methodology to other forms of media or cultural contexts

REFERENCES

[1] F. Zhang, Q. Tang, J. Chen and N. Han, “China public emotion analysis under normalization of COVID-19 epidemic: Using Sina Weibo,” *Front. Psychol*., vol. 3, no. 1066628, pp. 1-14, Jan. 2023, doi: 10.3389/fpsyg.2022.1066628.

[2] M. A. AlGhamdi and M. A. Khan, “Intelligent Analysis of Arabic Tweets for Detection of

Suspicious Messages," *Arab. J. Sci. Eng.*, vol. 45, pp. 6021–6032, Mar. 2020, https://doi.org/10.1007/s13369-020-04447-0.

[3] U. Upadhyay, A. Kumar, G. Sharma, S. Sharma, V. Arya,P. K. Panigrahi and B. B. Gupta, "A systematic data-driven approach for targeted marketing in enterprise information system," *Enter. Inf. Syst.*, vol. 18, no. 8, pp, 1058–1080, May 2024, https://doi.org/10.1080/17517575.2024.2356770.

[4] D. Bertero and P. Fung, "Predicting humor response in dialogues from TV sitcoms," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5780-5784, March 2016, doi: 10.1109/ICASSP.2016.7472785.

[5] M. Bedi, S. Kumar, M. S. Akhtar and T. Chakraborty, "Multi-modal sarcasm detection and humor classification in code-mixed conversations," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1363-1375, May 2021, https://doi.org/10.48550/arXiv.2105.09984.

[6] U. Chauhan, and A. Shah, "Topic modeling using latent Dirichlet allocation: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 7, pp. 1-35, Sept. 2021, https://doi.org/10.1145/3462478.

[7] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexicon-based methods for sentiment analysis," *Computat. Linguist.*, vol. 37, no. 2, pp. 267-307, Sept. 2010, https://doi.org/10.1162/COLI_a_00049.

[8] M. K. Hasan, S. Lee, W. Rahman, A. Zadeh, R. Mihalcea, L. P. Morency and E. Hoque, "Humor knowledge enriched transformer for understanding multimodal humor," *Proceedings of the AAAI conference on Artificial Intelligence*, vol. 35, No. 14, pp. 12972-12980, May 2021. https://doi.org/10.1609/aaai.v35i14.17534.

[9] A. Kotelnikova, D. Paschenko, K. Bochenina and E. Kotelnikov, "Lexicon-based methods vs. BERT for text sentiment analysis," *International Conference on Analysis of Images, Social Networks and Texts*, pp. 71-83, Dec. 2021, https://doi.org/10.48550/arXiv.2111.10097.

[10] R. Catelli, S. Pelosi and M. Esposito, "Lexicon-based vs. Bert-based sentiment analysis: A comparative study in Italian," Electronics, vol. 11, no. 374, pp. 1-20, Jan. 2022, https://doi.org/10.3390/electronics11030374.

[11] L. Augustyniak, T. Kajdanowicz, P. Szymański, W. Tuligłowicz, P. Kazienko, R. Alhajj and B. Szymanski, "Simpler is better? Lexicon-based ensemble sentiment classification beats supervised methods," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pp. 924-929, Aug. 2014, doi: 10.1109/ASONAM.2014.6921696.

[12] S. M. Mohammad and P. D. Turney, "NRC emotion lexicon," *National Research Council, Canada*, vol. 2, pp. 1-234, Nov. 2013. https://doi.org/10.4224/21270984.

[13] J. Miazga and T. Hachaj, "Evaluation of most popular sentiment lexicons coverage on various datasets," *2nd International Conference on Sensors, Signal and Image Processing,* pp. 86-90, Oct. 2019, https://doi.org/10.1145/3365245.3365251.

[14] M. Al-Shabi, "Evaluating the performance of the most important Lexicons used to Sentiment analysis and opinions Mining," *IJCSNS*, vol. 20, no. 1, pp. 51-75, Jan. 2020.

[15] Anjum, Aftab, and Nikolaus Lieberum. "Exploring Humor in Natural Language Processing: A Comprehensive Review of JOKER Tasks at CLEF Symposium 2023." *CLEF (Working Notes)* (2023): 1828-1837.

[16] Wu, Jiaming, Hongfei Lin, Liang Yang, and Bo Xu. "MUMOR: A Multimodal Dataset for Humor Detection in Conversations" in *CCF International Conference on Natural Language Processing and Chinese Computing*, pp. 619-627. 2021.

[17] Merlo, Lucia I., Berta Chulvi, Reynier Ortega, and Paolo Rosso. "When humour hurts: linguistic features to foster explainability." (2023).

[18] Chen, Yuyan, Yichen Yuan, Panjun Liu, Dayiheng Liu, Qinghao Guan, Mengfei Guo, Haiming Peng, Bang Liu, Zhixu Li, and Yanghua Xiao. "Talk funny! a large-scale humor response dataset with chain-of-humor interpretation" in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 16, pp. 17826-17834. 2024.

[19] Annamoradnejad, Issa, and Gohar Zoghi. "ColBERT: Using BERT sentence embedding in parallel neural networks for computational humor." *Expert Systems with Applications 249* (2024): 123685.

[20] S. Kublik and S. Saboo, "GPT-3: The Ultimate Guide To Building NLP Products With OpenAI API", *Packt Publishing Ltd.*, Birmingham, UK, 2022.

[21] R. M. Forte, "Mastering predictive analytics with R: Master the Craft of Predictive Modeling by Developing Strategy, Intuition, and a Solid Foundation in Essential Concepts", *Packt Publishing Ltd.*, Birmingham, UK, 2015.
[22] https://www.kaggle.com/datasets/gokulrajkmv/family-guy-dialog
[23] https://www.tvsubtitles.net/tvshow-62.html
[24] https://www.kaggle.com/datasets/eswarreddy12/family-guy-dialogues-with-various-lexicon-ratings/data
[25] https://www.kaggle.com/datasets/prashant111/the-simpsons-dataset
[26] https://www.kaggle.com/datasets/thedevastator/seinfeld-dialogue-dataset
[27] https://www.kaggle.com/datasets/josephvm/futurama-seasons-16-transcripts
[28] https://www.kaggle.com/datasets/tovarischsukhov/southparklines