Opinion Mining of Movie Reviews*

Nilesh Popli[1], Badal Nagpal[2], Narayan Chaturvedi[3], and Ankur Choudhary[4]

[1] Department of Computer Science and Engineering, Graphic Era (Deemed) University, Dehradun, Uttarakhand, India
nileshpopli23@gmail.com
[2] Department of Computer Science and Engineering, Graphic Era (Deemed) University, Dehradun, Uttarakhand, India
badalnagpal.2001@gmail.com
[3] Department of Computer Science and Engineering, Graphic Era (Deemed) University, Dehradun, Uttarakhand, India
narayanchaturvedi@gmail.com
[4] Department of Computer Science and Engineering, Graphic Era (Deemed) University, Dehradun, Uttarakhand, India
--@gmail.com

Abstract. The growing popularity of social media sites has generated a massive amount of data that attracted researchers, decision-makers, and companies to investigate people's opinions and thoughts in various fields. Opinion Mining is considered an emerging topic recently. Decision-makers, companies, and service providers as well-considered sentiment analysis as a valuable tool for improvement. In this paper, we aim to tackle the problem of sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis. Data used in this study are online movie reviews collected from IMDB. But our main aim is not just to classify reviews into positive, negative, neutral, but to use this, to analyse the pattern and the trend, which can help movie makers as well as the audience to get better movies in upcoming time.

Keywords: Movies · Reviews · Sentiments

1 Introduction

1.1 Sentiments

Sentiment analysis is the study of people's attitude, opinions, emotions towards a specific entity from written language It is an active research area which is blooming with the fastest pace as compared to any other aspect in the Natural Language Processing (NLP). Sentiment analysis help people by providing the basic gist of the data that the user might want to go through, it saves time for the user to go through the whole data and then put up their decision. A user goes through a collection of data due to many reasons like: going through data of movies to conclude which movie can be considered as the best one according to

* Supported by organization x.

one's preferences, going through data of stocks, where to invest in, how much to invest in, going through data of songs, which could be the best one for the current year, it's even being used in politics, like people tend to give their reviews about different candidates of a specific political party, analysis on the basis of their data can depict who might be a winning candidate for an upcoming election, analysis of people's reviews on a particular coffee shop and its respective products, etc. Sentiment analysis has been used at almost every place which requires the users to make decision based on some data, and this analysis helps the organization to come up with different products according to what the users want, or what they tend to expect from them. It also helps the companies to determine strategies on how to improve their products.

1.2 Approach

There are mainly 3 levels of Sentiment Analysis: document level, sentence level, aspect level, most researchers have been focusing on assigning sentiments to documents (document level). There are broadly 2 categories of doing Sentiment Analysis: Dictionary based and based on ML Classification. Dictionary based Sentiment Analysis is entirely based on dictionaries, sentiments are found by checking the values for specific words already present in a dictionary, whereas for ML Classification the values for the words are predicted using past values of related words. Here, our model performs coarse grained as well as fine grained Analysis determining the sentiment orientation and our Final Objective was to classify movies based on their genres, popularity, production house, Profit-Loss, movies that are awarded, etc. We visualized the scores of the reviews based on various classes of topics like genre, popularity, production house, profit-loss as positive, negative, and neutral in the form of a count plot using pyplot so that it's easy to read and understand for all the users and they need not go through the whole data. We considered the neutral sentiments too apart from the positive and negative to provide users with a better decision-making approach. Every dataset of different movies has a huge number of reviews and mapping them with other attributes of the movies. The scores were calculated using Afinn, which already has some specific words defined in it with a specific score. It fetches the whole sentence (review here) and based on the words gives it a sentiment. Here, we have done Sentiment Analysis of movies based on their reviews. Two methods were used to merge the datasets. The reason why we used reviews for sentiment analysis is because calculating scores based on reviews is the best way to analyse whether the movie is good or bad according to perspective of the audience.

1.3 Data

People tend to waste their time searching on what movies to watch after they finish one, they keep on surfing through different sites (including Rotten Tomatoes, IMDB), so to make it feasible for our users to decide we came up with this model. For doing Sentiment Analysis, the first and foremost objective is to find the appropriate and the accurate dataset, which is the main part of making any sentiment model. Both the types of Datasets we took is from the website of IEEE. We took 2 types of Datasets: 1) Multiple Datasets for each movie with their movie title, reviews with userID, ratings, date of review, as given by specific users (1150 datasets). 2) Multiple Datasets for each genre, containing the movie name, year of release, duration of movie, release date, overall rating, and URLs of reviews of the respective movies (17 datasets). 3) A single Dataset consisting of the Popularity, Production House of each movie, Profit-loss encountered by each movie, etc. 4) A single Dataset depicting which movies are awarded and which are not. Dictionary based novel approach has been proposed to Analyze the sentiments of movie viewers. The Sentiment Analysis Model that we have made is by far the best work in this aspect of sentence level sentiment analysis of movies using reviews with the help of Afinn, and no one has specifically done this kind of work yet to the best of my knowledge. Our method produces most accurate sentiment results/values as compared to the of document level sentiment analysis. Our first and foremost priority was to make our results accurate, so there was no point of any error, or if it exists, a very minute error, that could be neglected, our second priority was to make our data easy to read for our users, so they don't have to waste much of their time in reading it out, or understanding it, and to keep our results properly managed.

1.4 Contributions

Our major contributions are: 1) Data sets based on different movies and its attributes along with their movies. 3) Developed a Lexicon Based approach to analyze movie reviews.

2 Literature Review

V.K. Singh et al. [13] came up with the conclusion that most of the Researchers tend to have their focus on determination of only sentiment orientation (positive versus negative), here they used SentiWordNet based scheme. The algorithms performed were Naive Bayes, K-Nearest Neighbour, Random Forest. Only the naive Bayes classifier came up with the best results.

As per the research done by Tun Thura Thet et al. [14], sentiment analysis was performed at the clause level using a linguistic approach. In this approach both domain-specific lexicon and a generic opinion lexicon were used, derived from SentiWordNet, to assign a sentiment score to each word in a sentence beforehand and analysis was done on the basis of those sentiments.

Atiqur Rahman et al. [9] tested their model with five kinds of machine learning classifiers to analyse the data. The used classifiers were Bernoulli Naive Bayes (BNB), Decision Tree (DE), Support Vector Machine (SVM), Maximum Entropy (ME), as well as Multinomial Naive Bayes (MNB). According to their results, MNB achieved better accuracy, and precision. And still both lacked having more accurate results coz of not testing the model with more algorithms apart from these.

Ankita Gandhi et al. [4] used Multimodal Sentiment Analysis in their Research. Multimodal sentiment analysis is a subset of traditional text-based sentiment analysis that includes other improvements such as speech and visual features along with the text. Visual features give a visual depiction and can describe something more efficiently than a whole long list of written words and may could be highly utilized to predict the appropriate sentiment according to the data.

Houshmand Shirani-Mehr et al. [12] started their research with a preliminary investigation of the dataset. Naive Bayes baseline classifier was implemented on the dataset, then different deep learning techniques were applied on it, and their performances were analysed. Various types of neural networks were used in this research.

In the research paper as given, according to Palak Baid et al. [2], they used Machine Learning Algorithms like Naïve Bayes, K-Nearest Neighbour (KNN), Random Forest. According to the Naive Bayes classifier, the presence of a particular feature in a class is not related to the presence of any other feature. The KNN method is used to find a predefined number of training samples closest in distance to the new point and predict the label from these. Random Forests method constructs several decision trees during training time. To classify new case, it sends the new case to each of the trees. Each tree performs classification and outputs a class.

According to the research of Nehal Mohamed Ali et al. [1] four deep neural networks were introduced with sentiment analysis: MLP, CNN, LSTM in addition to a hybrid model CNN-LSTM and results of these four models were compared with the results of SVM and Naïve Bayes, also the results of the proposed models were compared to the results recorded by applying RNTN to an English movies reviews dataset.

Minhoe Hur et al., 2016 [5] proposed a system to predict Box-office collection based on Sentiments of movie review. They have used Viewer opinions are used as input variables in addition to predictors and three machine learning-based algorithms (artificial neural network, regression tree, and support vector regression) were used to get non-linear relationship between the box-office and its collection predictors.

Aurangzeb Khan, 2011 [6] proposed a rule-based technique in which SentiWordNet is used to obtain more accuracy than a pure lexicon-based technique for sentiment analysis for customer reviews and software reviews. The proposed system has 97.8 percent of accuracy at the feedback level and 86 percent of accuracy at the sentence level.

Mudinas and Zhang, 2012 [7] proposed a hybrid technique which gives better performance than the lexicon and almost performs like leaning based technique. Hybrid Techniques are stable as lexicon technique and performance as machine learning based techniques. The system has an overall accuracy of 82.3 percent.

Lei Zhang et al., 2010 [16] proposed a ranking and extracting product features in opinion documents algorithm. Initially, they have reviews of users and it was difficult to determine by the machine to differentiate between positive reviews and negative reviews. They used the associated rule mining technique for extracting product features.

Seven Rill et al., 2014 [10] proposed an application "PoliTwi" which shows Early detection of emerging political topics on Twitter and the impact on concept-level sentiment analysis. In this Paper twitter, hashtags are used to determine the results of the election in the USA even before "Google Trends". Twitter API is used to collect data and Analyze using sentiment analyzing an algorithm.

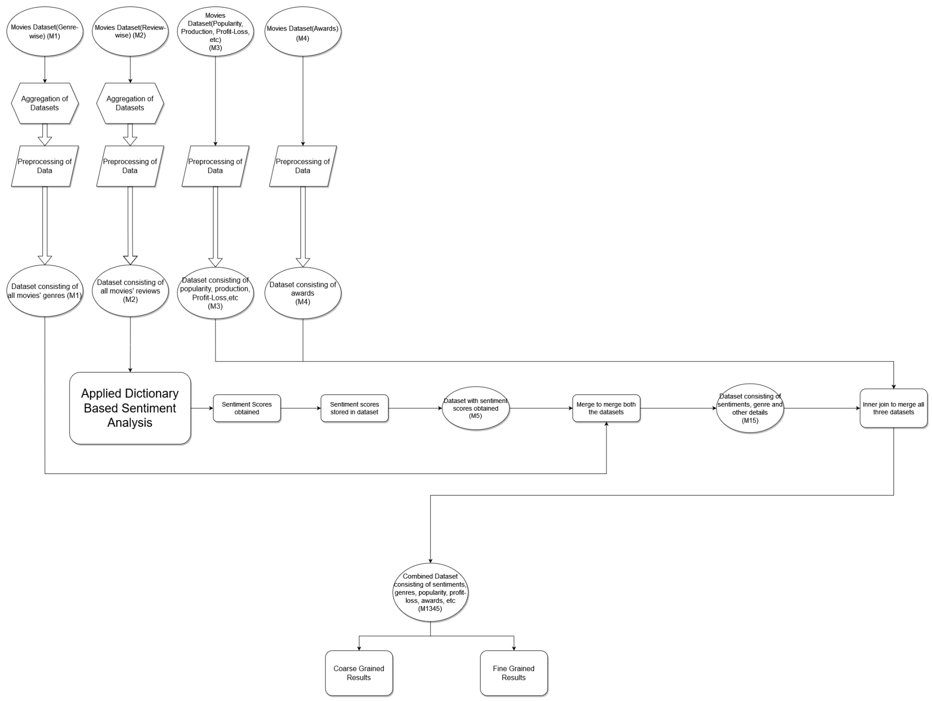
Martin Wöllmer et al., 2013 [15] proposed a technique to analyze sentiments in Audio – Video context of a YouTube Movie. They used Metacritic database to get user reviews as input. They evaluated the knowledge-based approach, applying data-based approach in an in-domain setting as well as in a cross-domain setting.

Giuseppe Di Fabbrizio et al. 2013 [3] proposed aspect rating distributions and language modeling which used for summarizing online product and service reviews. They used a novel approach for extracting multi-document summarization for textual data that considers aspect rating distributions and language modeling as summarization features.

Rafeeque Pandarachalil et al., 2014 [8] proposed a method for Twitter sentiment analysis using an unsupervised learning approach. They determined the Polarity of tweets is evaluated by using three sentiment lexicons-SenticNet, SentiWordNet, and SentislangNet. They used parallel python framework to implement this method.

Chirag Sangani 2013 [11] proposed a method for analyzing user sentiments towards apps through their review comments and ratings can be economically profitable to app developers. They propose a system that provides a list of reviews for each topic that represents user opinions towards that topic and a many-to-many relation portraying from reviews to topics of interest.

3 Methodology



Methodology Flowchart

3.1 Datasets Collection

Data Collection is the first stage in the Sentiment Analysis. The required datasets are collected from various sources. Reviews per movie dataset, genre wise movies dataset, taken from IMDB, metadata of movies dataset, and Oscar award dataset is collected and the further used for analysis. Reviews of 1000+ movies are taken into account, where reviews ae given movie wise in individual datasets, and there are at least 400 reviews for each movie. These datasets are merged and mapped with other datasets. We have many details for movie, like director name, budget, language, country, release date, and many more.

3.2 Datasets Cleaning and Pre-processing

Reviews contains a lot of opinions about the movies which are expressed in different ways by different users. The raw data having polarity is highly susceptible to inconsistency and redundancy, therefore it is pre-processed before applying algorithm.

a) Dropping columns which are not required from the datasets. This is done to reduce the complexity. b) Converting review attribute to lower case. It is important for an effective analysis. c) Tokenization: We simply split the text into

```
Data columns (total 24 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   adult                  8754 non-null   bool
 1   belongs_to_collection  1264 non-null   object
 2   budget                 8754 non-null   int64
 3   genres                 8693 non-null   object
 4   homepage               2559 non-null   object
 5   id                     8754 non-null   int64
 6   imdb_id                8751 non-null   object
 7   original_language      8753 non-null   object
 8   original_title         8754 non-null   object
 9   overview               8678 non-null   object
 10  popularity             8754 non-null   float64
 11  poster_path            8740 non-null   object
 12  production_companies   7958 non-null   object
 13  production_countries   8472 non-null   object
 14  release_date           8747 non-null   object
 15  revenue                8754 non-null   int64
 16  runtime                8732 non-null   float64
 17  spoken_languages       8551 non-null   object
 18  status                 8753 non-null   object
 19  tagline                4847 non-null   object
 20  title                  8754 non-null   object
 21  video                  8754 non-null   bool
 22  vote_average           8754 non-null   float64
 23  vote_count             8754 non-null   int64
```

Fig. 1. Movies Metadata

```
Data columns (total 7 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   username  311 non-null    object
 1   rating    311 non-null    object
 2   helpful   311 non-null    int64
 3   total     311 non-null    int64
 4   date      311 non-null    object
 5   title     311 non-null    object
 6   review    311 non-null    object
```

Fig. 2. Reviews Per Movie

```
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   name          100 non-null    object
 1   year          100 non-null    int64
 2   movie_rated   100 non-null    object
 3   run_length    100 non-null    object
 4   genres        100 non-null    object
 5   release_date  100 non-null    object
 6   rating        100 non-null    float64
 7   num_raters    100 non-null    int64
 8   num_reviews   100 non-null    int64
 9   review_url    100 non-null    object
```

Fig. 3. Movies Per Genre

```
Data columns (total 7 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Decade         4671 non-null   int64
 1   year_ceremony  4671 non-null   int64
 2   ceremony       4671 non-null   int64
 3   category       4671 non-null   object
 4   name           4671 non-null   object
 5   film           4564 non-null   object
 6   winner         4671 non-null   bool
```

Fig. 4. Oscar Movies

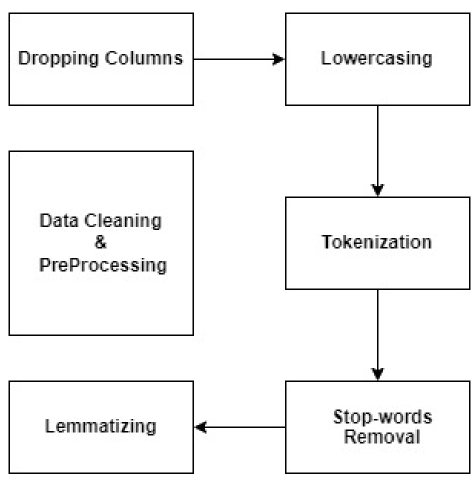


Fig. 5. Pre-processing

tokens. In other words, the string is converted into a list, where each element corresponds to a word. The module nltk provides tokenize function to tokenize the text. d) Stop word Removal: Stop words that don't affect the meaning of the text are removed (for example and, or, still etc). e) Lemmatizing: We want to convert a word to its base form. For example, playing, plays and play can seem different to the computer, but they are the same thing. We need to produce the root forms of these words. The nltk module plays a relevant role this time too. It provides the WordNetLemmatizer function, which looks for the lemmas of the words using a database, called WordNet.

3.3 Sentiment Classification

Lexicon based method uses sentiment dictionary with opinion words and match them with the data to determine polarity. Lexicon-based approaches mainly rely on a sentiment lexicon, i.e., a collection of known and precompiled sentiment terms, phrases and even idioms, developed for traditional genres of communication, such as the Opinion Finder lexicon.

The Affin Lexicon is perhaps one of the simplest and most popular lexicons that can be used extensively for sentiment analysis. They assign sentiment scores from -5 to +5 to the opinion words, and find the sentiments for the text, if [polarity]>0, sentiment is positive if [polarity]<0, sentiment is negative. if [polarity]=0, sentiment is neutral. Dictionary-based: It is based on the usage of terms (seeds) that are usually collected and annotated manually. This set grows by searching the synonyms and antonyms of a dictionary. An example of that dictionary is WordNet, which is used to develop a thesaurus called SentiWordNet.

3.4 Plotting Graphs

Visualising the data is one of the major parts of opinion mining projects. All the datasets are merged, and the results are plotted. Catplot (categorical plots) is being plotted between a particular attribute with respect to Sentiments, and then is done same for all the attributes. We do this for fine grain classification. Percentage of count is used as a factor in these graphs. This is done to analyse the data

3.5 Result Analysis

The plotted graphs are used to analyse the fine grain course for coarse grain classification, and then it is concluded.

4 Result

As we can see, these graphs show us the users' opinion on movies, with various aspects, and therefore we can find the pattern in it. In all, we can analyse whether user likes the movie or not or is neutral, based on the sentiment score, from the reviews. With these different aspects, we can see how the movie is being taken by the audience, which matters the most for the movie makers. Also, we are concerned with the percentage or ratio of positive and negative sentiments, rather than the count, as it could vary, since for a popular movie, the sentiments count will be more, but that need not necessarily means that it was better received by the audience.

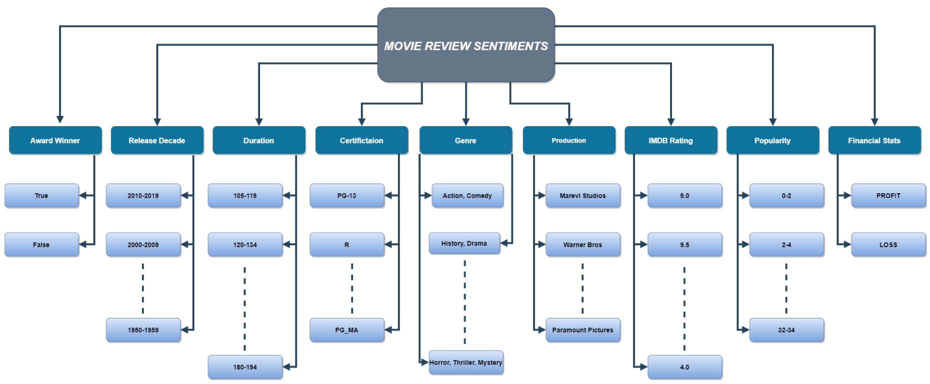
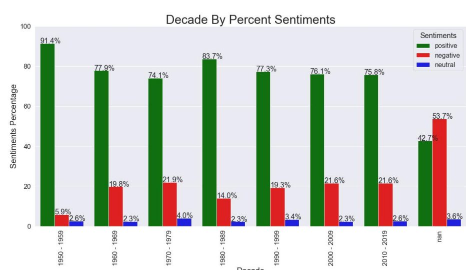


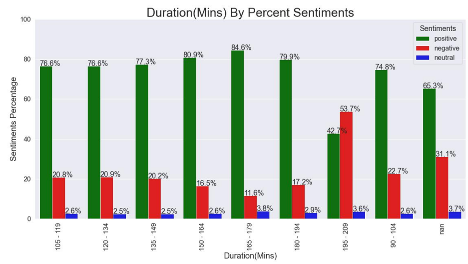
Fig. 6. Result Flowchart

4.1 Decade Wise Sentiments



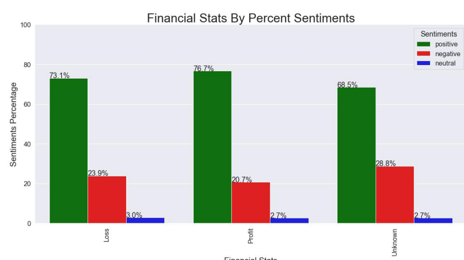
We can analyse that in the decade, 1950-1959, the positive sentiment percentage is max, which shows us that old movies are well received by the audience, while in recent decade, i.e., 2010-2019, negative sentiment percentage is high, which shows that the latest movies are not that good as per the audience.

4.2 Duration Wise Sentiments



We can see by the graph, that the movies with duration 150-164 165-179 minutes are very well received, and movie with duration 195-209 are not very well received, therefore makers must try to keep the duration near about 2.5 hours.

4.3 Financial Stats Wise Sentiments



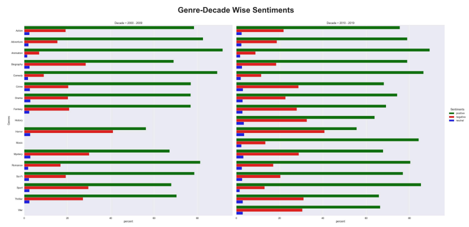
We can see that the movie which earned profit at the box office, were also liked by the audience, while for the movies the performed bad at the box office, were not liked by the audience, which explains why it earned less.

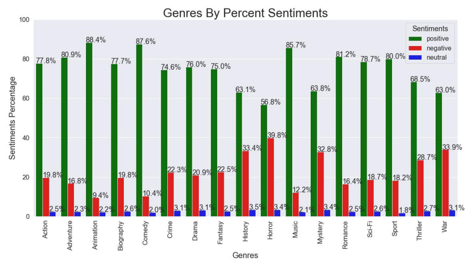
4.4 Genre-Decade Wise Sentiments

Analysing the pattern in both the decades, we can see that, comedy films, musical films are now more liked by the audience, while romantic movies are little less liked by the audience.

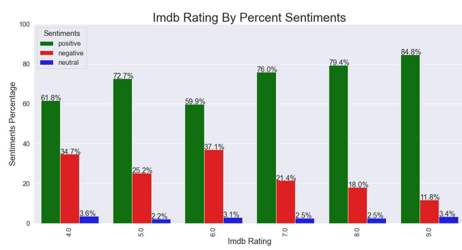
4.5 Genre Wise Sentiments

Analysing the pattern, we can see that, comedy films, action films are liked, while romantic movies are little less liked by the audience.





4.6 IMDB Rating Wise Sentiments



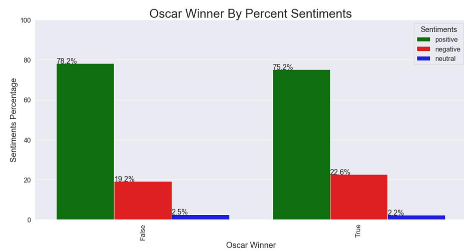
We can see that the movie with max IMDB ratings, has received most positive feedback, which shows that the IMDB ratings affect the viewers mindset, and matters to the public. But the movies with rating 5, also shows a decent positive graph, which shows, that not all audience are affected by the critics.

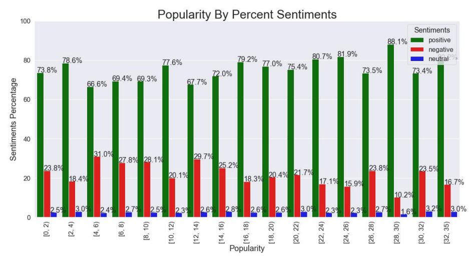
4.7 Oscar Wise Sentiments

The graph shows us that, movies which have either not won an Oscar or have not been nominated has received positive reviews more than movies which won, which shows that award winning movies may or may not be liked by audience.

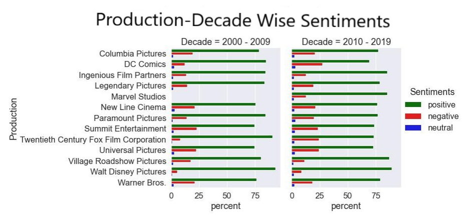
4.8 Popularity Wise Sentiments

We can see that, movies with high popularity score, are wither well received by the audience, or received pretty bad, which should also have been the case, as the film generally gets popular if it is either really good or really bad.





4.9 Production-Decade Wise Sentiments



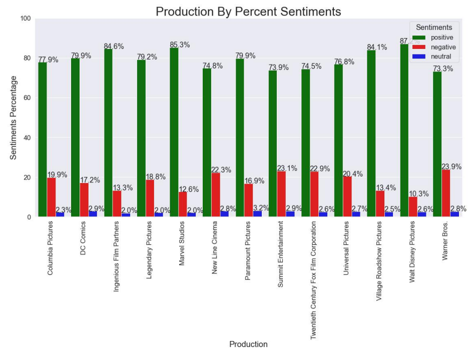
The graph shows us that Dc comics production house has declination of positive feedback, while its the exact opposite as marvel studios came recently, and dethroned dc comics.

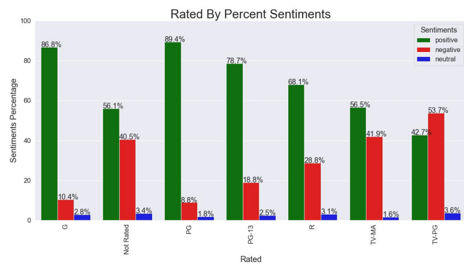
4.10 Production Wise Sentiments

By the graph we can see that Walt Disney and Marvel studious have the highest positive feedback as it is relatable to every age group and are undoubtedly, the best production houses currently.

4.11 Certification Wise Sentiments

We can see that movies with certification PG 13 have the most positive percentage, while TV-PG has received most negative feedback, which shows that in TV people don't usually prefer PG films, while its the exact opposite in the case of cinemas, as most people like to watch movies with family.





5 Conclusion

Sentiment analysis of movie reviews has great potential for understanding and predicting audience reactions to movies, as it aims in providing insights into aspects of movies which are most relevant for audiences. We also experimented with different text pre-processing techniques to improve the result. We analysed a dataset of movie reviews from IMDb to classify them as positive, negative, or neutral.
Our findings can be useful for movie studios, critics, and audiences to make informed decisions about which movies to watch or recommend.

6 Bibliography

References

1. Ali, N.M., Abd El Hamid, M.M., Youssif, A.: Sentiment analysis for movies reviews dataset using deep learning models. International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol **9** (2019)
2. Baid, P., Gupta, A., Chaplot, N.: Sentiment analysis of movie reviews using machine learning techniques. International Journal of Computer Applications **179**(7), 45–49 (2017)
3. Di Fabbrizio, G., Aker, A., Gaizauskas, R.: Summarizing online reviews using aspect rating distributions and language modeling. IEEE Intelligent Systems **28**(3), 28–37 (2013). https://doi.org/10.1109/MIS.2013.36

4. Gandhi, A., Adhvaryu, K., Khanduja, V.: Multimodal sentiment analysis: Review, application domains and future directions. In: 2021 IEEE Pune Section International Conference (PuneCon). pp. 1–5 (2021). https://doi.org/10.1109/PuneCon52575.2021.9686504
5. Hur, M., Kang, P., Cho, S.: Box-office forecasting based on sentiments of movie reviews and independent subspace method. Information Sciences **372**, 608–624 (2016). https://doi.org/https://doi.org/10.1016/j.ins.2016.08.027, https://www.sciencedirect.com/science/article/pii/S0020025516306016
6. Khan, A., Baharudin, B., Khan, K.: Sentiment classification from online customer reviews using lexical contextual sentence structure. In: Mohamad Zain, J., Wan Mohd, W.M.b., El-Qawasmeh, E. (eds.) Software Engineering and Computer Systems. pp. 317–331. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)
7. Mudinas, A., Zhang, D., Levene, M.: Combining lexicon and learning based approaches for concept-level sentiment analysis. In: Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining. WISDOM '12, Association for Computing Machinery, New York, NY, USA (2012). https://doi.org/10.1145/2346676.2346681, https://doi.org/10.1145/2346676.2346681
8. Pandarachalil, R., Sendhilkumar, S., Mahalakshmi, G.: Twitter sentiment analysis for large-scale data: an unsupervised approach. Cognitive computation **7**(2), 254–262 (2015)
9. Rahman, A., Hossen, M.S.: Sentiment analysis on movie review data using machine learning approach. In: 2019 International Conference on Bangla Speech and Language Processing (ICBSLP). pp. 1–4 (2019). https://doi.org/10.1109/ICBSLP47725.2019.201470
10. Rill, S., Reinel, D., Scheidt, J., Zicari, R.V.: Politwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis. Knowledge-Based Systems **69**, 24–33 (2014). https://doi.org/https://doi.org/10.1016/j.knosys.2014.05.008, https://www.sciencedirect.com/science/article/pii/S0950705114001920
11. Sangani, C., Ananthanarayanan, S.: Sentiment analysis of app store reviews. Methodology **4**(1), 153–162 (2013)
12. Shirani-Mehr, H.: Applications of deep learning to sentiment analysis of movie reviews. In: Technical report. Stanford University (2014)
13. Singh, V., Piryani, R., Uddin, A., Waila, P.: Sentiment analysis of movie reviews and blog posts. In: 2013 3rd IEEE International Advance Computing Conference (IACC). pp. 893–898 (2013). https://doi.org/10.1109/IAdCC.2013.6514345
14. Thet, T.T., Na, J.C., Khoo, C.S., Shakthikumar, S.: Sentiment analysis of movie reviews on discussion boards using a linguistic approach. In: Proceedings of the 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion. p. 81–84. TSA '09, Association for Computing Machinery, New York, NY, USA (2009). https://doi.org/10.1145/1651461.1651476, https://doi.org/10.1145/1651461.1651476
15. Wöllmer, M., Weninger, F., Knaup, T., Schuller, B., Sun, C., Sagae, K., Morency, L.P.: Youtube movie reviews: Sentiment analysis in an audio-visual context. IEEE Intelligent Systems **28**(3), 46–53 (2013). https://doi.org/10.1109/MIS.2013.34
16. Zhang, L., Liu, B., Lim, S.H., O'Brien-Strain, E.: Extracting and ranking product features in opinion documents. In: Coling 2010: posters. pp. 1462–1470 (2010)