

GROUP PROJECT REPORT

ANALYTICS OF MANAGERIAL DECISION MAKING (IBM-322)

Streaming Sentiments: Analyzing Emotions on OTT Platforms



PROFESSOR

Prof. Sumit Kumar Yadav

MEMBERS

Priyanshu Arya (21324019)

Garima Mandli (21113048)

Anvita Deepak Prasad (20122010)

Roopa Pramodan (21324021)

Shoilyee Chaudhuri (21112102)



TABLE OF CONTENTS

1. Introduction
 - 1.1 Motivation
 - 1.2 Snapshot of data
2. Approach
3. Methodology
 - 3.1 Data Extraction
 - 3.2 Text Preprocessing
 - 3.3 Sentiment Analysis
 - 3.4 Linguistic Analysis
 - 3.5 Data Analysis
4. Results and insights
5. Conclusion
6. Reference
7. GitHub link of Project Work



1. Introduction

The way we view films and TV episodes has been completely transformed by Over-The-Top (OTT) platforms in an era where digital content consumption is the norm. A huge selection of content is available on services like Netflix, Amazon Prime Video, Hulu, and JioCinema, which appeal to a wide range of interests and inclinations. However, these platforms are more than simply entertainment; they are also a rich source of data that might reveal attitudes among the viewers.

1.1 Motivation

This initiative is driven by the growing recognition of the value of audience emotion and preference analysis in the context of digital content. Given the growing popularity of OTT services, more research is required to understand how consumers perceive the information they watch. We hope to get insightful information about audience reactions by examining the attitudes stated in articles about these platforms. This information will assist marketers, platform providers, and content creators better understand their target audiences and adjust their services. We collected data for 4 different platforms from articles available on the web. In short, our project is about making content experiences more personal, engaging, and relevant.

1.2 Snapshot of Data

Link to google document containing data:

https://docs.google.com/spreadsheets/d/168htVyxOSPWIYWLW75EvEuCEiYU8GJrAi/edit?usp=drive_link&ouid=114929682142028180378&rtpof=true&sd=true



The screenshot shows an Excel spreadsheet with a single column labeled 'URL_ID' and a header 'URL of Articles'. The spreadsheet contains 27 rows of URLs, mostly from business and technology news sources like Business Insider, The Guardian, and Wired. The interface includes standard Excel menus (File, Edit, View, Insert, Format, Data, Tools, Help) and a ribbon with various formatting options.

| URL_ID | URL of Articles |
|--------|---|
| 1 | https://www.businessinsider.com/netflix-strategy-big-budget-action-films-big-stars-2024-478b-1 |
| 2 | https://www.businessinsider.in/tech/news/how-netflix-is-changing-the-global-entertainment-industry/articleshow/85430648.cms |
| 3 | https://advancetian.com/opinion/2022/04/27/netflix-isnt-dying-its-already-dead |
| 4 | https://lancerglobal.com/insights/digital-transformation/the-evolution-of-netflix-from-renting-dvds-to-establishing-global-streaming-supremacy#:~:text=The%20evolution%20of%20Netflix%20 |
| 5 | https://www.theguardian.com/media/2022/feb/05/stream-big-how-netflix-changed-the-tv-landscape-in-10-years |
| 6 | https://www.sartanalytics.com/insights/original-series-rising-in-demand-are-netflix-originals-still-at-the-top/ |
| 7 | https://journals.sagepub.com/doi/full/10.1177/17596020211037259 |
| 8 | https://pioneersperspective.com/innovation/how-netflix-has-changed-the-global-entertainment-industry/ |
| 9 | https://quartz.com/insights/company-research/inside-netflix-innovation-originals-and-cultural-phenomena |
| 10 | https://www.wired.com/story/how-do-netflix-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like/#:~:text=Netflix%20uses%20machine%20learning%20and%20algorithm |
| 11 | https://www.wired.com/story/how-do-netflix-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like/ |
| 12 | https://dl.acm.org/doi/pdf/10.1145/2843988 |
| 13 | https://netflixtechblog.com/recommender-for-the-world-8da8c051b |
| 14 | https://towardsdatascience.com/deep-div-into-netflix-recommender-system-341806ae3b48 |
| 15 | https://bbc.org/2018/10/how-netflix-expanded-to-190-countries-in-7-years |
| 16 | https://press.farm/netflix-expansion-reed-hastings-global-strategy/ |
| 17 | https://www.weglot.com/blog/netflix-localization-strategy/ |
| 18 | https://www.accelingo.com/netflix-localization-strategy/ |
| 19 | https://www.saglobal.com/marketintelligence/news-insights/research/netflix-global-pricing-strategy-2022 |
| 20 | https://www.researchgate.net/publication/36939937_Analysis_of_Netflix%27s_Strategy_and_Innovation_in_a_Global_Context_Based_on_BCG_Matrix |
| 21 | https://www.forbes.com/sites/dereksaul/2024/04/18/netflix-reports-record-profits-as-subscriber-growth-tops-estimates/7ahc258ad48d22b |
| 22 | https://www.forbes.com/sites/dereksaul/2024/04/18/netflix-reports-record-profits-as-subscriber-growth-tops-estimates/7ahc258ad48d22b |
| 23 | https://www.forbes.com/sites/dereksaul/2024/04/18/netflix-reports-record-profits-as-subscriber-growth-tops-estimates/7ahc258ad48d22b |
| 24 | https://www.forbes.com/sites/dereksaul/2024/04/18/netflix-reports-record-profits-as-subscriber-growth-tops-estimates/7ahc258ad48d22b |
| 25 | https://www.forbes.com/sites/dereksaul/2024/04/18/netflix-reports-record-profits-as-subscriber-growth-tops-estimates/7ahc258ad48d22b |
| 26 | https://www.forbes.com/sites/dereksaul/2024/04/18/netflix-reports-record-profits-as-subscriber-growth-tops-estimates/7ahc258ad48d22b |
| 27 | https://www.forbes.com/sites/dereksaul/2024/04/18/netflix-reports-record-profits-as-subscriber-growth-tops-estimates/7ahc258ad48d22b |

2. Approach

In order to accomplish this, we have painstakingly collected a wide variety of articles from numerous sources that address a wide range of subjects pertaining to the content that is accessible on well-known OTT platforms like Netflix, Amazon Prime, Hulu, and JioCinema. These articles, which include opinions, news updates, reviews, and critiques, give readers a complete picture of how people feel about these platforms' material.

In order to ensure a wide range of viewpoints and opinions, our collection approach involves searching via a variety of internet venues, including blogs, social networking platforms, news portals, and entertainment websites. In order to provide a comprehensive grasp of audience sentiments, we tried to incorporate articles from both expert critics and general viewers.

3. Methodology

The project demonstrates the application of web scraping, NLP techniques, and sentiment analysis for extracting and analyzing textual data from online sources. By computing linguistic features and sentiment scores, valuable insights can be derived to understand the textual content's characteristics and sentiment orientation.

3.1 Data Extraction

The data extraction phase utilizes the Scrapy library for web crawling and scraping. This Python framework allows automated retrieval of HTML content from URLs stored in an Excel file. BeautifulSoup is then employed to parse the HTML structure and extract relevant text data from webpages. The extracted



data is saved into individual text files for subsequent preprocessing and analysis.

3.2 Text Preprocessing

Text preprocessing is a critical step aimed at refining raw textual data for further analysis. In this project, spaCy's English language model is utilized for removing stop words, such as articles and prepositions, from the extracted text. Additionally, unwanted data is eliminated to ensure that only relevant information is retained for subsequent analysis tasks.

3.2.1 Natural Language Processing

In the realm of Natural Language Processing (NLP), several techniques are employed to process and analyze text data effectively. In our project, we utilize key NLP techniques to preprocess and analyze the text extracted from articles about Over-The-Top (OTT) platforms.

Tokenization

Tokenization involves breaking down the text into individual words or tokens. This process is essential for further analysis as it provides a structured representation of the text data. We use tokenization to segment the text into meaningful units, enabling subsequent processing steps such as sentiment analysis and word frequency analysis.

Stop Word Removal

Stop words are common words that occur frequently in the text but typically do not carry significant meaning or contribute to the context of the text. Examples include "the," "is," "and," etc. Stop word removal involves filtering out these non-essential words from the text data. By removing stop words, we can focus on the more meaningful content of the text, improving the accuracy of subsequent analyses.

3.3 Sentiment Analysis

In the sentiment analysis phase of our project, we employed a robust pre-trained model called the CardiffNLP Twitter RoBERTa model. The key objective of sentiment analysis is to evaluate the sentiment polarity of the extracted text data, which includes assessing whether the text expresses positive, negative, or neutral sentiments.



The CardiffNLP Twitter RoBERTa model assigns sentiment scores to the text, providing insights into the emotional tone and attitude conveyed by the content. These sentiment scores indicate the degree of positivity, negativity, and neutrality present within the text. By analyzing these scores, we can identify sentiment trends within the data, such as prevalent positive or negative sentiments associated with specific topics or platforms.

This sentiment analysis process enables us to gain a deeper understanding of audience perceptions and attitudes towards the content available on Over-The-Top (OTT) platforms. By uncovering sentiment trends, we can identify areas of strength and areas for improvement, guiding strategic decision-making for marketers, platform providers, and content creators. Overall, sentiment analysis serves as a valuable tool for deciphering audience sentiment and informing content optimization strategies to enhance audience engagement and satisfaction.

3.4 Linguistic Analysis

Linguistic analysis focuses on extracting linguistic features and characteristics from the text data. This includes counting the total number of words, identifying complex words, calculating syllables per word, and determining average word length. Additionally, metrics such as sentence length, percentage of complex words, and the Fog Index are computed to provide insights into the linguistic complexity of the text.

3.5 Data Analysis

Data analysis encompasses the aggregation and interpretation of insights derived from the processed text data. Various metrics, including sentiment scores, linguistic features, and text statistics, are analyzed to uncover patterns, trends, and relationships within the data. This analysis aids in gaining a deeper understanding of the textual content and facilitates informed decision-making.

4. Results and Insights

We will examine sentiment ratings from various OTT platforms to find trends and patterns in audience responses. Do people who watch Netflix originals tend to feel better about them than they do about other platforms? Are there



certain content categories or kinds that garner more favorable reviews? Through our investigation, we want to provide answers to some of these issues.

5.Conclusion

Our initiative aims to shed light on the nuanced landscape of audience attitudes on over-the-top platforms. Through extensive article analysis and the application of sentiment analysis algorithms, our goal is to deliver insightful information that may guide content creation, marketing plans, and platform improvements. In the end, a more involved and fulfilling viewing experience for everyone can result from an awareness of the audience's emotions.

6. References

<https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/>

<https://www.ibm.com/topics/sentiment-analysis>

All the article links are in the input file of the dataset

7. Github link of Project work

<https://github.com/YASTREAMER/IBM322>