

NETFLIX OR DISNEY+?

PREDICTIVE MODELING FOR STREAMING SERVICE CLASSIFICATION



TABLE OF CONTENTS

01

INTRODUCTION

02

PROCESS

03

EXPLORATORY
DATA ANALYSIS (EDA)

04

MODELLING
& INSIGHTS

05

RECOMMENDATIONS

06

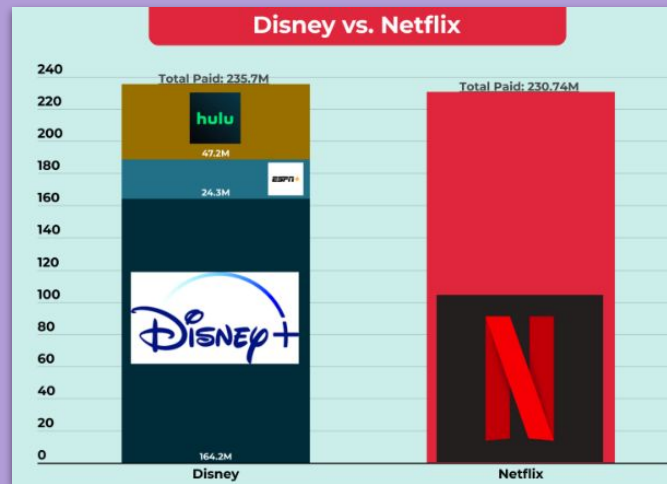
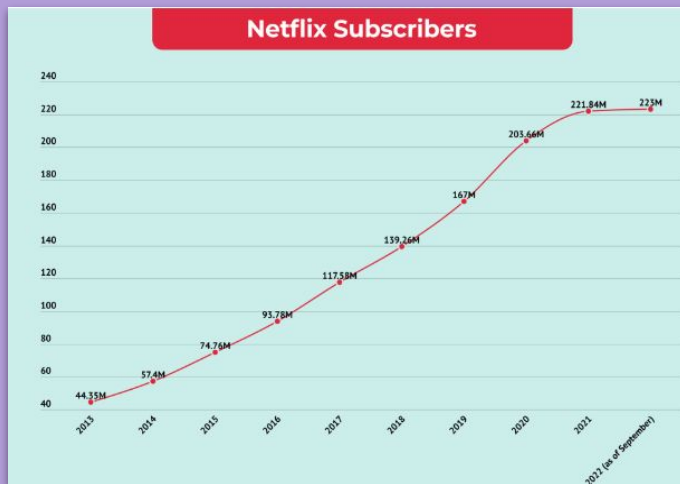
LIMITATIONS

01

THE PROBLEM

FORBES > MONEY

Disney Surpasses Netflix Subscriber Count: What Does That Mean For Investors?



NETFLIX KEEPS LOSING!

Disney Overtakes Netflix, Again. Who Will Win the Streaming Wars?

By Jennifer Saibil – Nov 13, 2022 at 11:15AM

NYSE: DIS

Walt Disney



Market Cap

\$148B

Today's Change

↓ (-0.19%) -\$0.15

Current Price

\$79.90

Price as of September 27, 2023, 4:00 p.m. ET

FINANCE · DISNEY

Netflix's bad year just got worse as Disney+ passes it in streaming subscriptions

BY SOPHIE MELLOR

August 11, 2022 at 7:33 PM GMT+8



Netflix just lost \$18 billion in value, oof



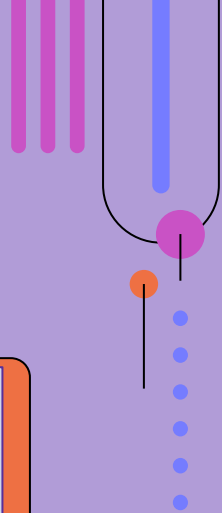
Richard Lee Breslin



Published 15:57, 21 July 2023 BST

| Last updated 15:57, 21 July 2023 BST

OUR TEAM



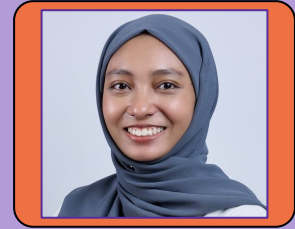
RAJ



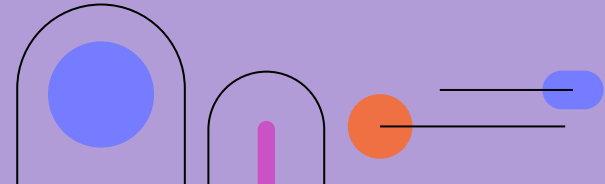
YEOW WEN



MAIMUNAH



LIYENA



OPTIMIZING NETFLIX'S ONLINE PRESENCE

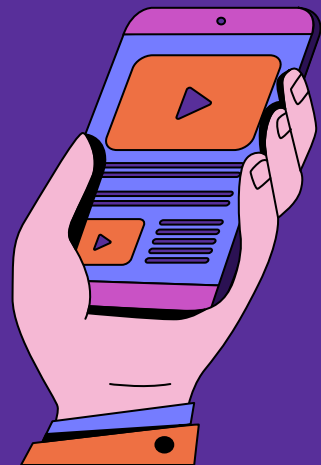
Who: We are a team of **data science consultants** hired by the **Netflix marketing team**.

Problem Statement: The **Netflix marketing team** seeks a model to analyze and ensure their marketing campaign **aligns with keywords and themes closely associated with Disney+**.

Goal: Achieve a **minimum of 90% F1 score**, aiming to strategically **redirect** online search traffic **towards Netflix**.

Why: This strategic maneuver will **help Netflix solidify its position** in the market and **attract more viewers** to their platform.

How: We're tasked with building a model to assess if Netflix's campaign aligns with Disney+ semantics, targeting a 90% F1 Score.



METRICS AND TARGET

PRECISION

How many of the words that we labeled as Netflix and Disney Plus are **actually correct**?

RECALL

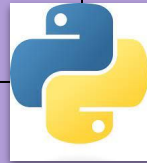
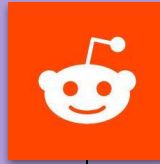
Of all the words that are Netflix and Disney Plus, how many of those were **correctly predicted**?

F1 SCORE: 90% - - - - **Imbalanced Dataset**

The F1 score measures the **balance** between precision and recall.

DATA SOURCES

REDDIT



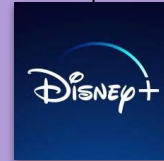
PYTHON

PRAW library
for interacting with the
Reddit API



NETFLIX

[https://www.reddit.com/
r/netflix](https://www.reddit.com/r/netflix)



DISNEY +

[https://www.reddit.com/
r/disneyplus](https://www.reddit.com/r/disneyplus)

02

PROCESS

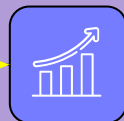
DATA
SCRAPING



DATA CLEANING
PREPROCESSING



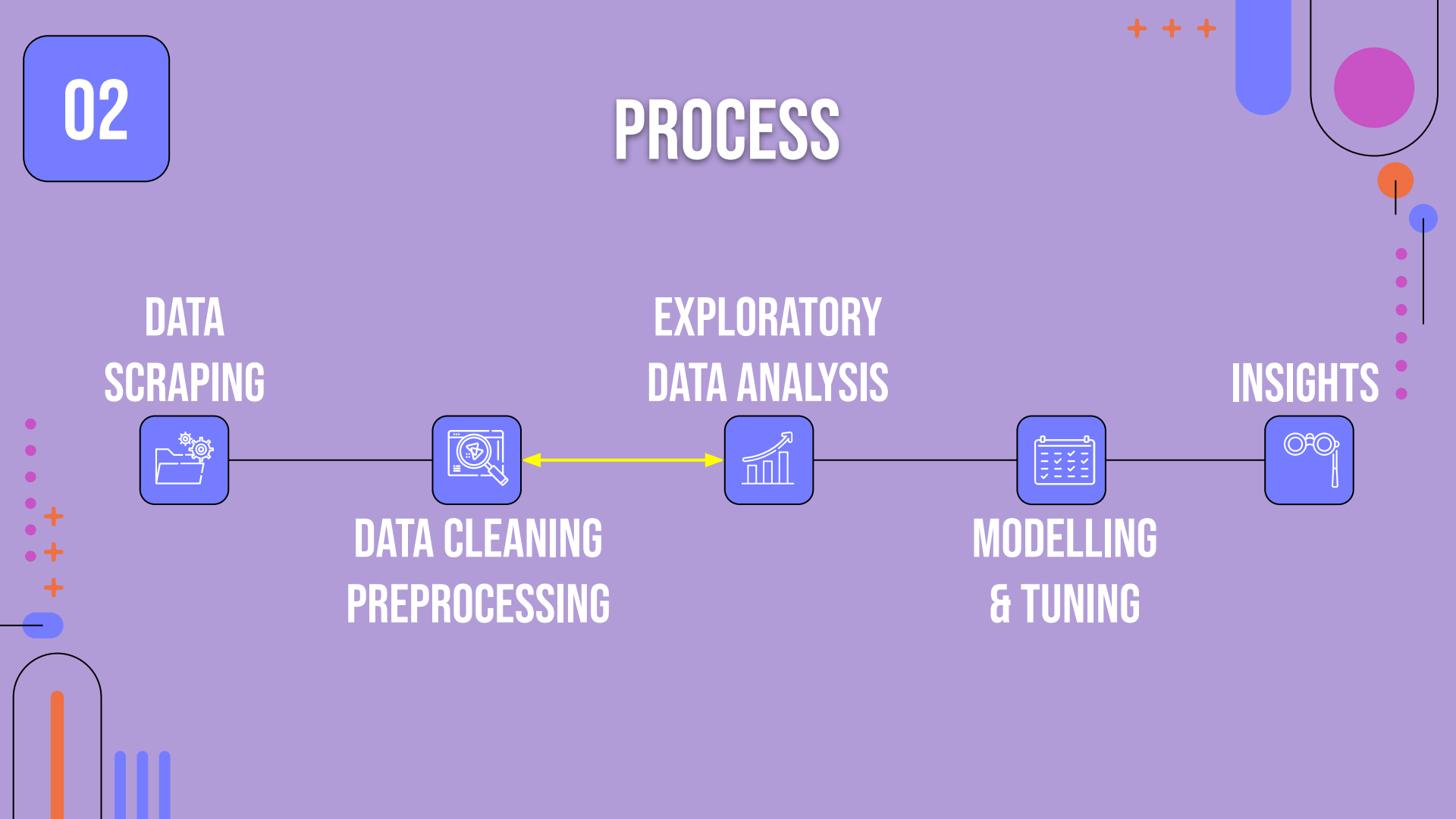
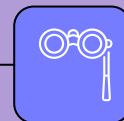
EXPLORATORY
DATA ANALYSIS



MODELLING
& TUNING



INSIGHTS





r/netflix

1893

unique Netflix posts

23 JUN '11 TO 26 SEP '23

period of posts

1329

missing rows



r/DisneyPlus

1803

unique DisneyPlus posts

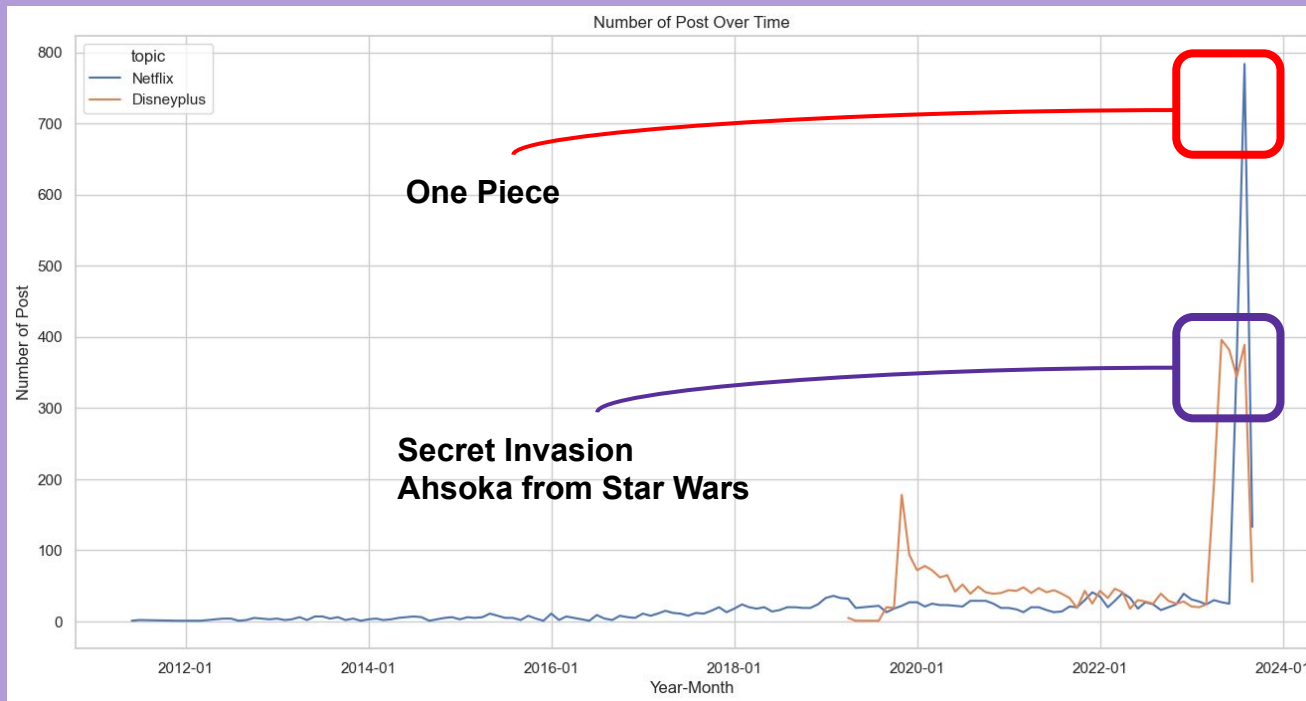
2 APR '19 TO 26 SEP '23

period of posts

1889

missing rows

NUMBER OF POSTS OVER TIME



- Post for Netflix are available earlier than Disneyplus
- Disneyplus launched on Nov 2019

DATA CLEANING - LEMMATIZATION

RAW TEXT

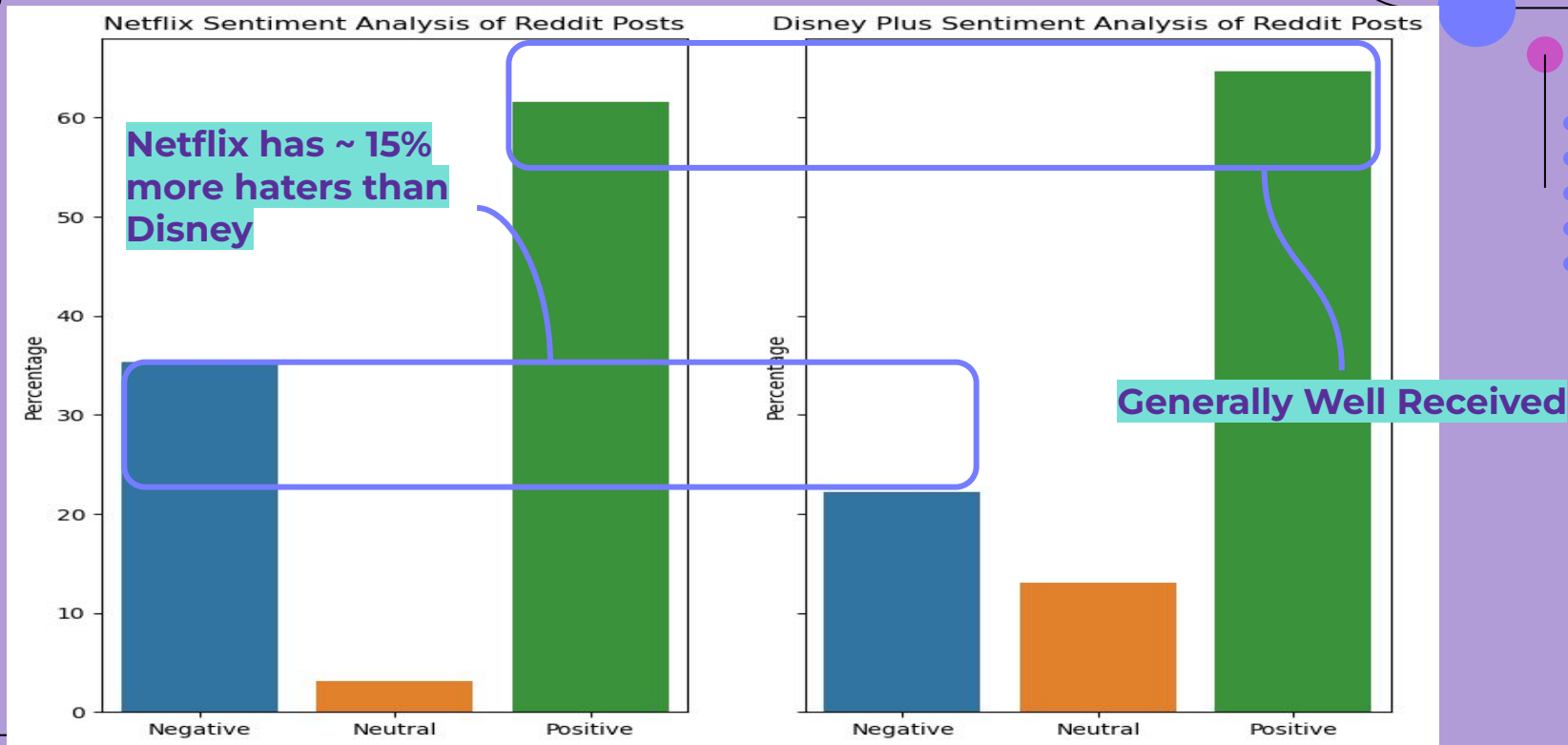
/r/Netflix Discord Server We are
pleased to announce we have
affiliated with
<https://discord.gg/Netflix> which
will be the subreddit Discord
server for the Netflix subreddit!
\n\nFeel free to join the server
and talk about everything Netflix
related, **including shows** on
Netflix as well :).



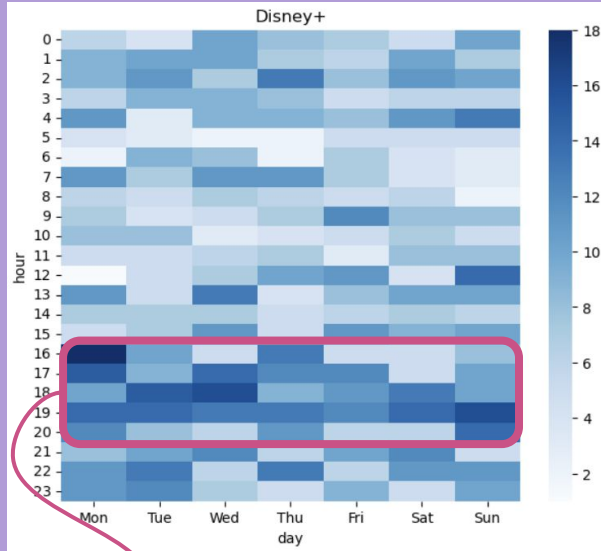
PROCESSED TEXT

r netflix discord server pleased
announce **affiliate** subreddit
discord server netflix subreddit
feel free join server talk
everything netflix related
include show netflix well

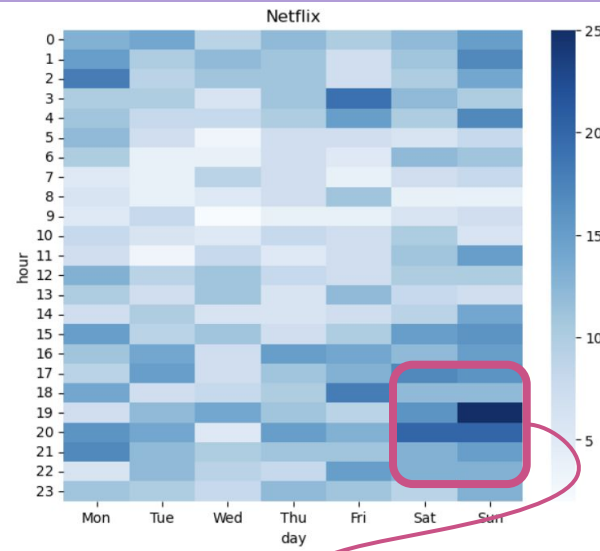
SENTIMENT ANALYSIS



WHEN DO THEY POST?

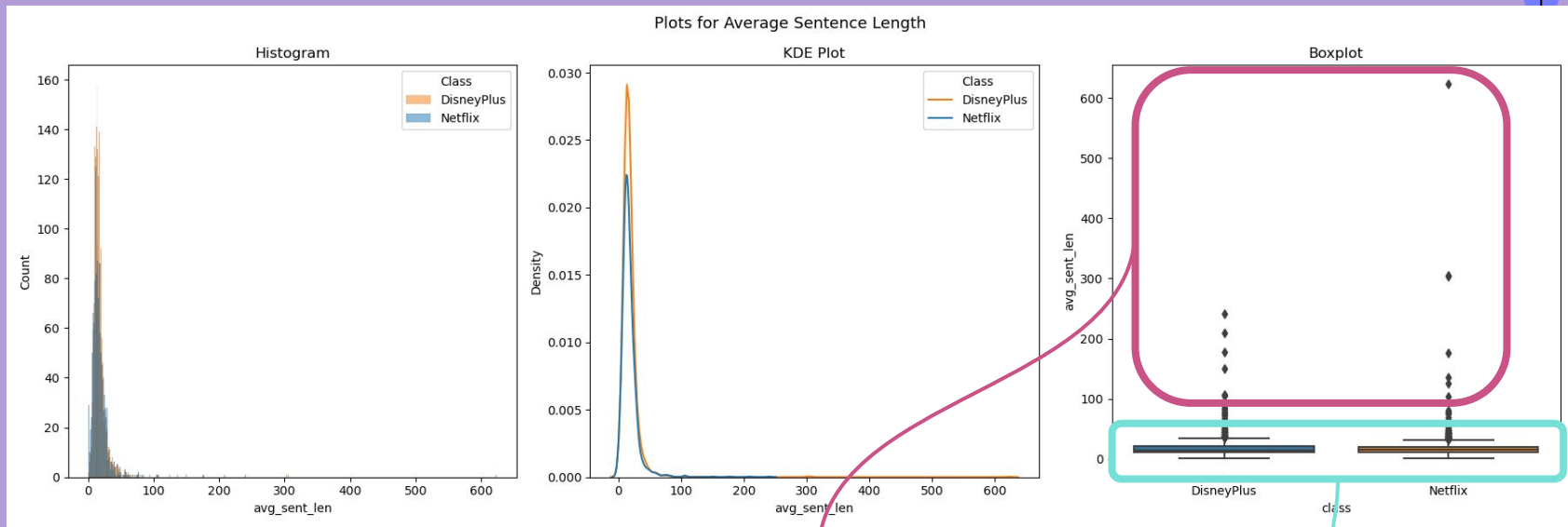


- Late afternoon/Early Evening
- Consistent



- Late afternoon/Early Evening
- Weekends only

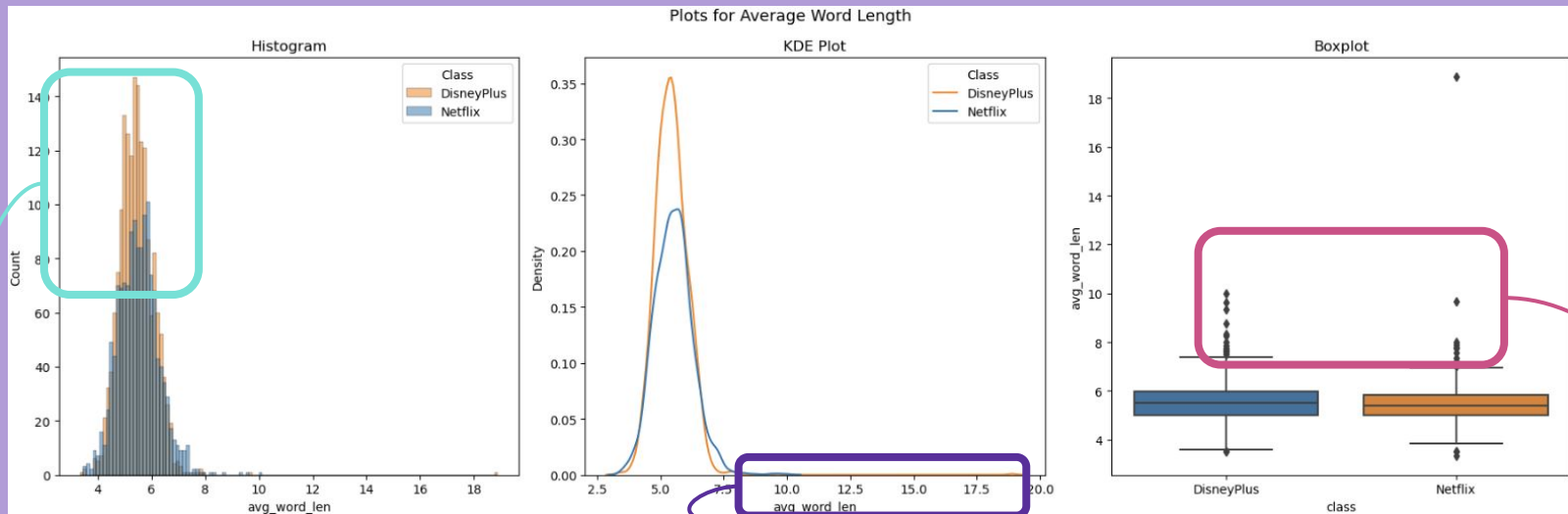
HOW LONG ARE USERS' COMMENTS?



- In depth discussions
- In depth reviews
- Long-winded comments

- Median: 15-20 words

WORD LENGTH ANALYSIS



- DisneyPlus posts use more words in general

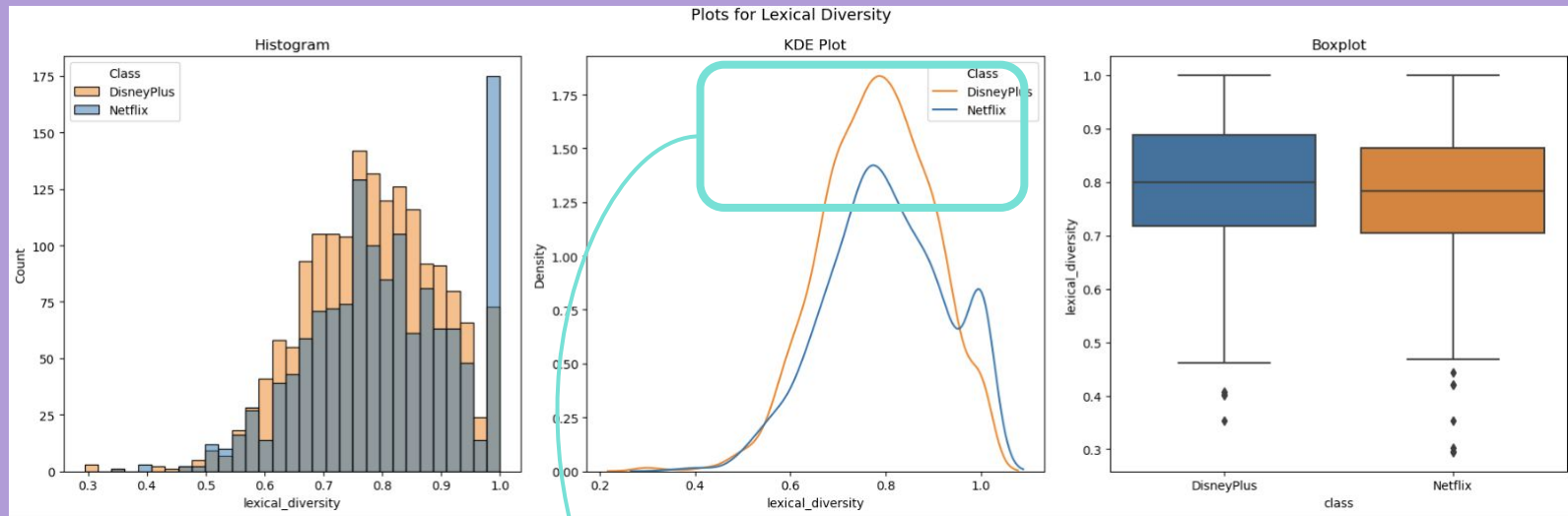
- Longer tail: Occasional use of longer words for DisneyPlus
- Composite terms merged into a single term

- Longer or Complex words
- Users posting without spaces

PITFALL OF AUTOMATED TEXT CLEANING

alltext	cleaned_alltext
All posts regarding tech support belong here.\n\nExamples of tech support questions are: \n\n*\nHow do I cancel?\n* Why does the app crash on my Fire Stick/Roku/Apple TV?\n* Why don't the subtitles work correctly?\n* I am being overcharged for my subscription.\n\nBrowse other tech support posts [here] (https://old.reddit.com/r/DisneyPlus/search?q=tech+support&restrict_sr=on&sort=relevance&t=all). This is the Weekly Tech Support Thread	post regard tech support belong example tech support question I cancel app crash stickrokuapple tv subtitle work correctly I overcharge subscription browse tech support post weekly tech support thread

VOCABULARY RICHNESS ACROSS POSTS



Disney+ users use more
different words

↑ lexical diversity
↑ richer vocabulary

netflix series cancel new series upload netflix
star war upload new disney plus

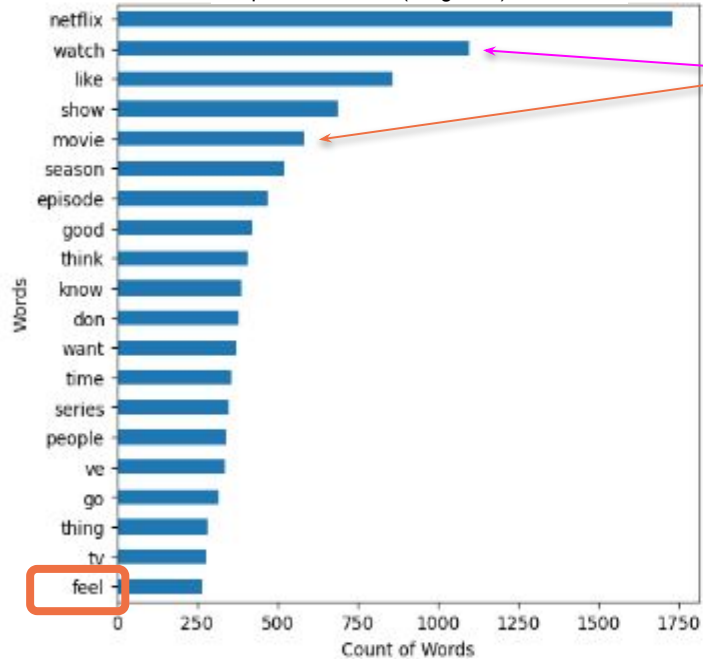
$$5 / 7 = 0.71$$

$$6 / 6 = 1$$

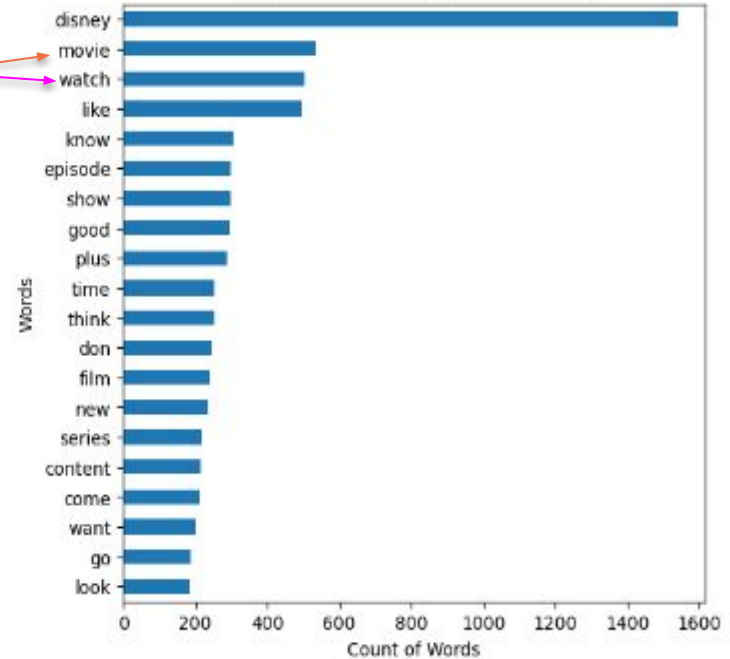


UNIGRAMS

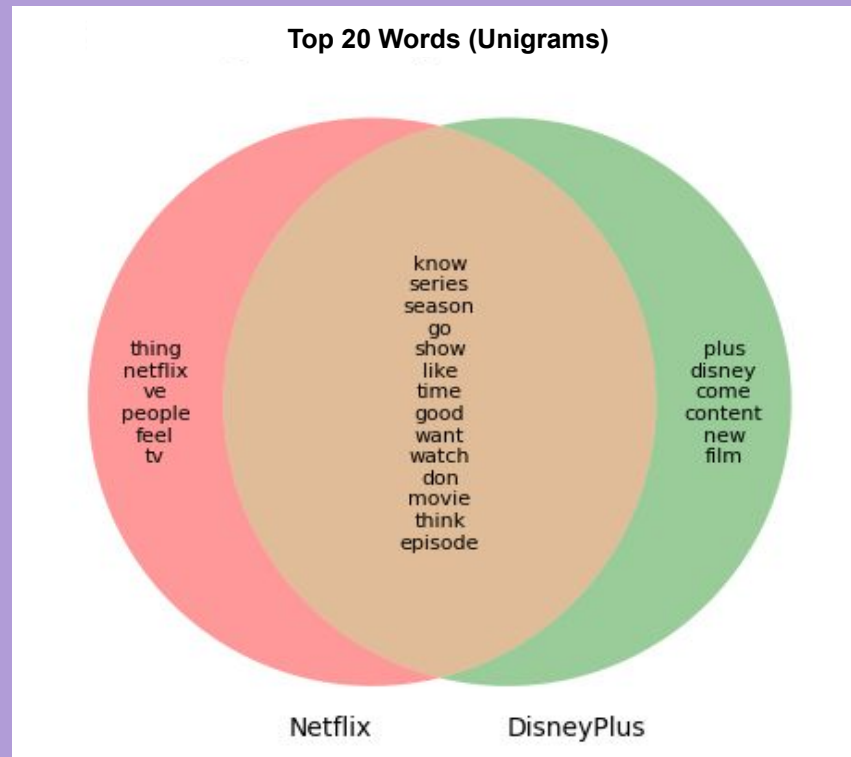
Top Netflix Words (Unigrams)



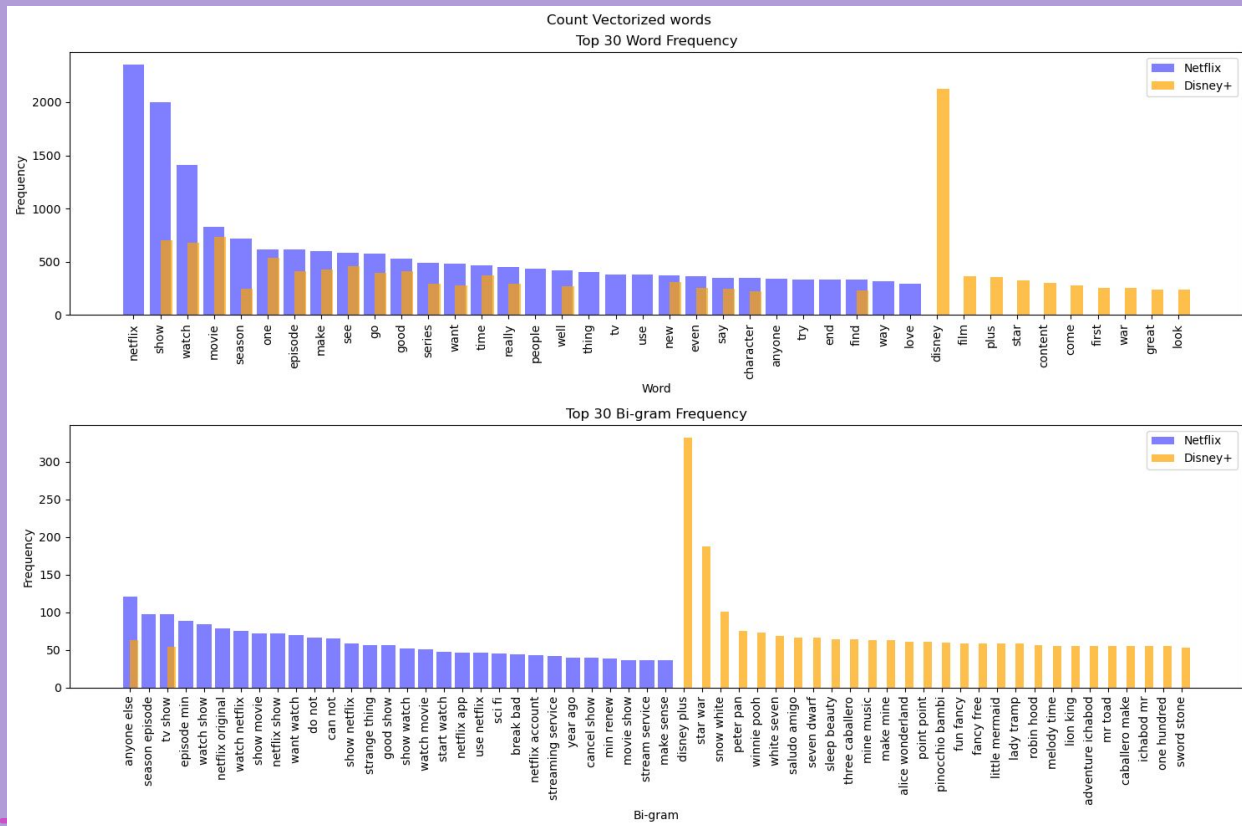
Top DisneyPlus Words (Unigrams)



WORD WARS: SHARED & EXCLUSIVE TERMS

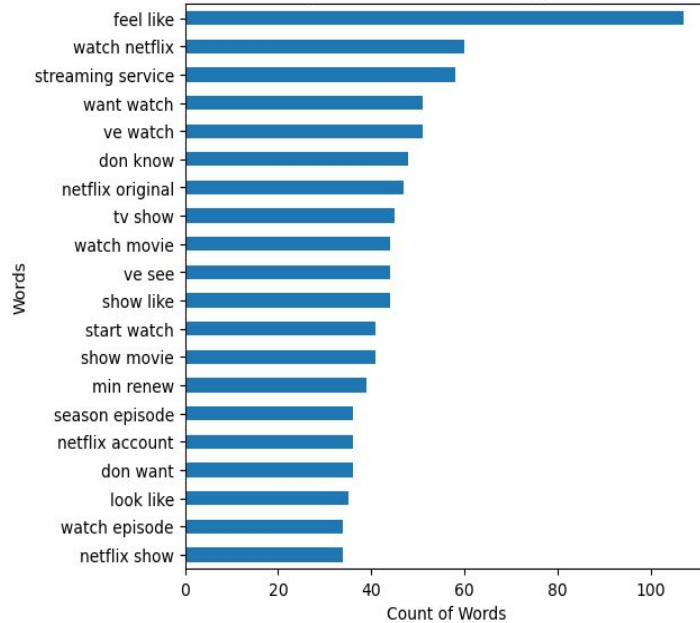


WORD WARS: SHARED & EXCLUSIVE TERMS

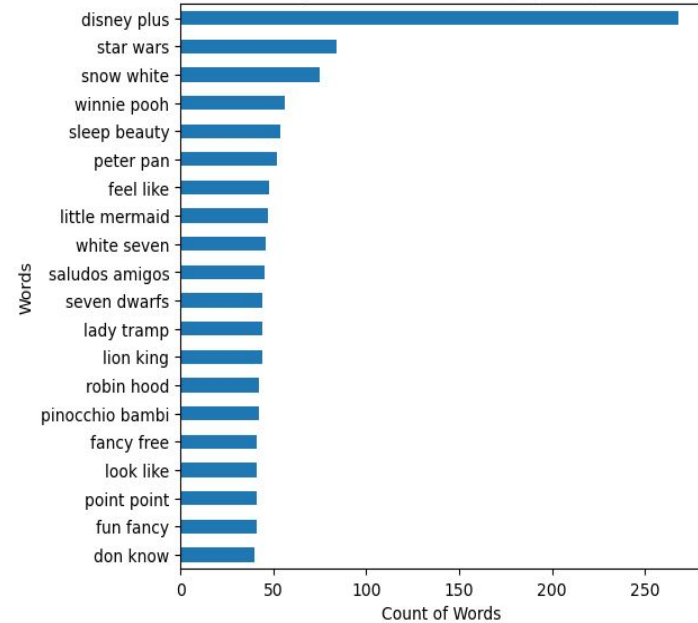


BIGRAMS

Top Netflix Words (Bi-grams)

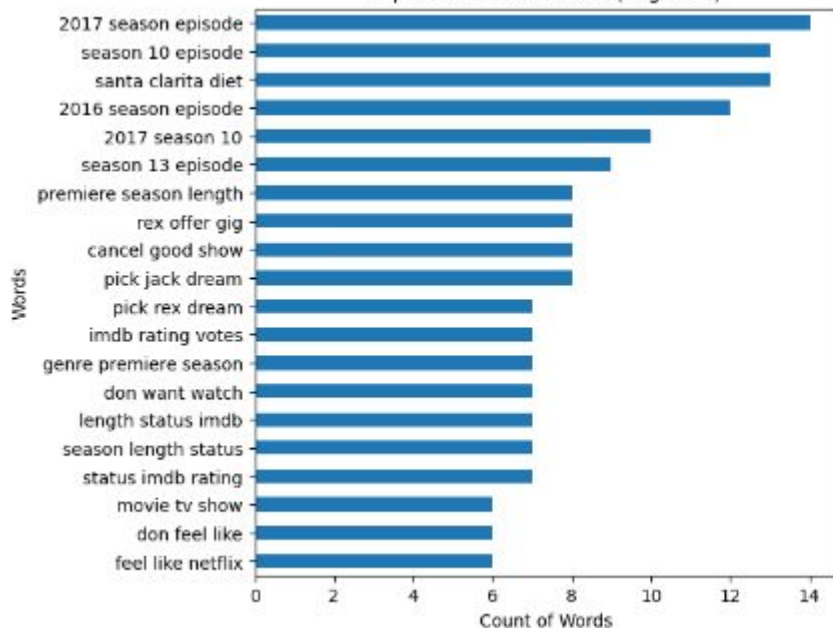


Top DisneyPlus Words (Bi-grams)

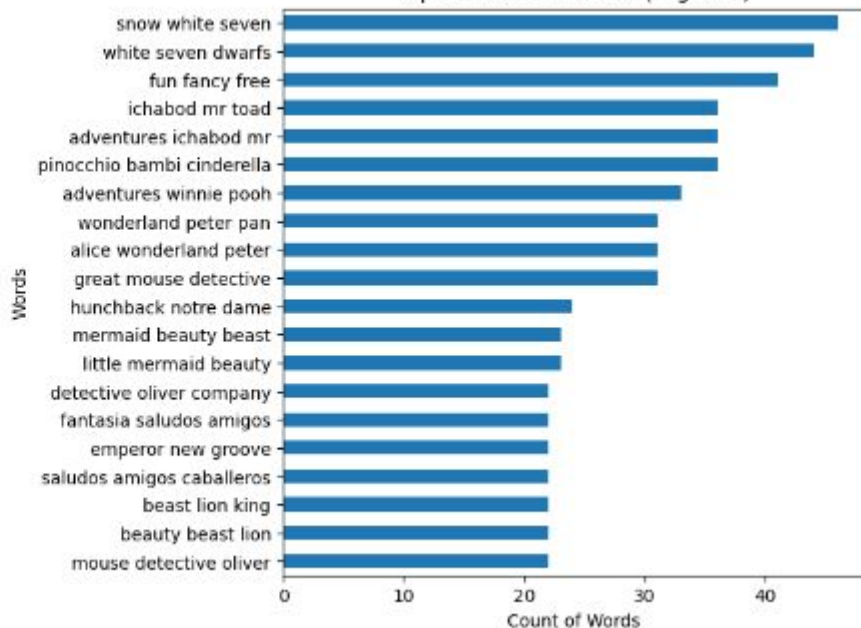


TRIGRAMS

Top NF Common words (Trigrams)



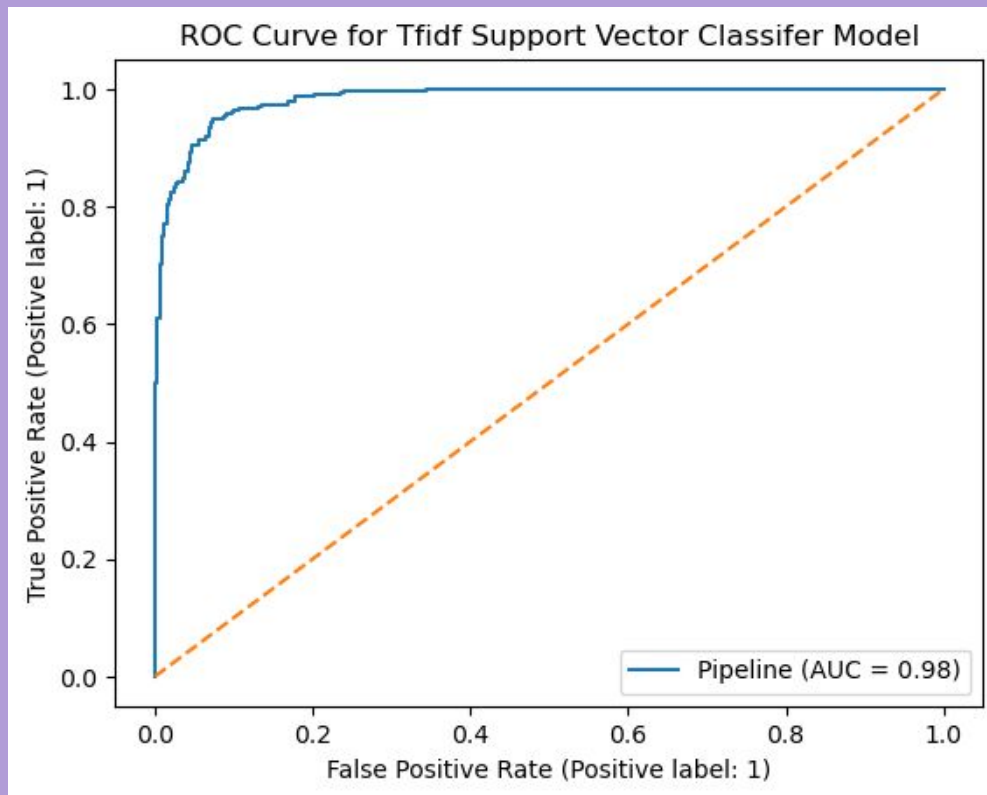
Top DP Common words (Trigrams)



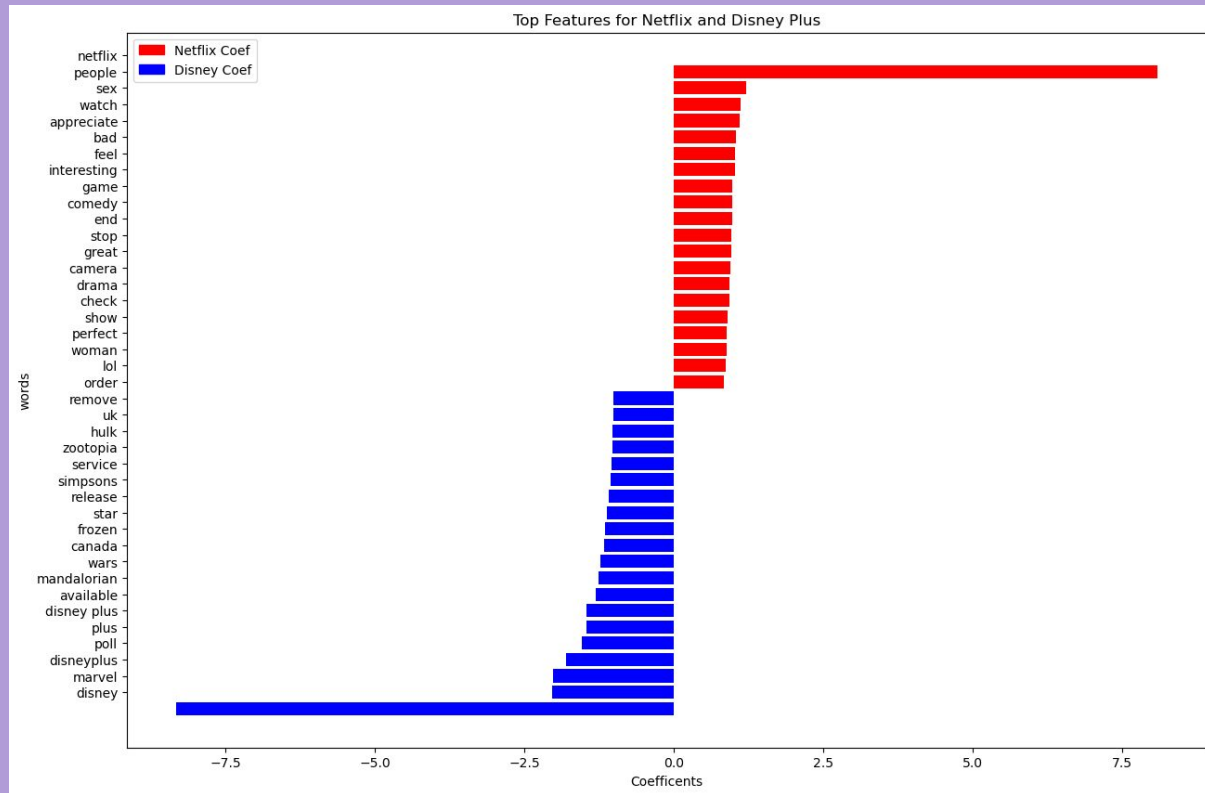
MODELLING & TUNING

VECTORIZER	MODEL	F1 (TRAIN)	F1 (TEST)
CountVectorizer	K-Nearest Neighbors Classifier (Baseline)	1	0.78
CountVectorizer	Naïve Bayes	0.95	0.91
TF-IDF Vectorizer	Naïve Bayes	1	0.87
CountVectorizer	Random Forest	0.93	0.90
TF-IDF Vectorizer	Random Forest	0.93	0.89
CountVectorizer	Support Vector Classifier (SVC)	0.97	0.90
TF-IDF Vectorizer	Support Vector Classifier (Chosen)	1	0.92
TF-IDF Vectorizer	K-Nearest Neighbors Classifier	1	0.88
CountVectorizer	Logistic Regression	1	0.9036
TF-IDF Vectorizer	Logistic Regression	1	0.92

ROC CURVE



INSIGHTS



RECOMMENDATIONS

1. **Leverage on Popular Disney Themes:**
Develop content in trending genres like sci-fi and superheroes to cater to DisneyPlus fans seeking variety.
2. **Enhance User Experience:**
Prioritize platform usability and innovative features to outshine competitors and retain customers.
3. **Content Marketing:**
Spotlight unique Netflix content to position the platform as either a complement or superior alternative to DisneyPlus.

LIMITATIONS

- **Dynamic Content Landscape**

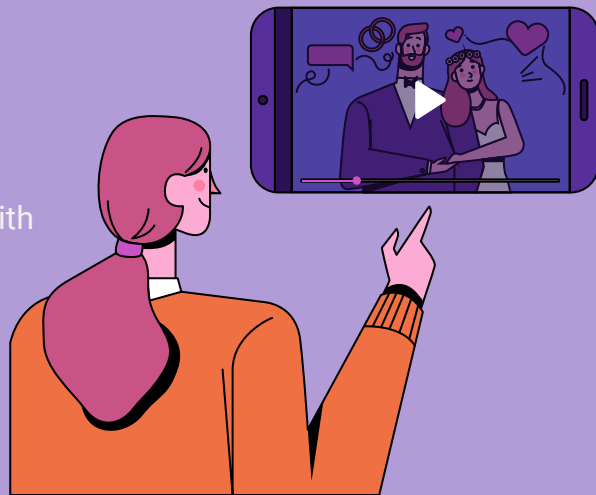
The content and associated keywords for both platforms are constantly evolving.

- **Feedback Loop Concerns**

Using Disney+ keywords on Netflix and retraining the model may blur semantic distinctions between the two platforms over time.

- **Semantic Nuances**

Text classification models can identify keywords but might struggle with understanding context or sentiment.



FURTHER IMPROVEMENTS

- **Data Expansion:** Collect more data over time to improve the model's robustness
- **Data Imbalance:** Implement algorithm such as SMOTE to handle increasing data imbalance over time.
- **Model Exploration:** Try more advanced models like deep learning for text classification.
- **Ensemble Methods:** Combine predictions from multiple models to improve metrics.

REFERENCES

- [Yahoo!news, 2019. Eye-popping chart shows how much original content Netflix is creating.](#)
- <https://fortune.com/2022/08/11/netflix-bad-year-just-got-worse-as-disney-passes-it-in-streaming-subscribers/>
- <https://www.fool.com/investing/2022/11/13/disney-overtakes-netflix-again-who-will-win-the-streaming-war/>
- <https://www.gamingbible.com/news/tv-and-film/netflix-just-lost-18-billion-in-value-949478-20230721>

THANKS!

Do you have any questions?



CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, infographics & images by Freepik

Please keep this slide for attribution

