

1. INTRODUCTION

1.1. Overview

Under the current scenario, social distancing measures have forced theatres to limit the number of audiences or even shut down and that encouraged people to stay at home, accelerating the increase in OTT platform subscriptions. Therefore, there is the need to analyse different OTT platforms and provide useful information for people who are not able to decide which platform fits them best.

Our solution is a web-app called **Find Your Binge** which will help our users to figure out which OTT platform is best suited for them. Our web-app will provide comparisons of the top OTT platforms visually so that the users can make some sense out of these statistics. It also provides an in-depth analysis of these OTTs and will tell them how people feel about these platforms. Along with that our app also provides with the latest trending content and updates from these streaming platforms. We have added a chatbot created using **IBM Watson Assistant**, that can help the user discover content over different OTT Platforms.

We have acquired different datasets related to OTTs from various credible sources to form a cumulative dataset. We have used various strategies to clean and process this data to make it comparable to each other. Then, we have converged the different data points into common features of the dataset to perform **Exploratory Data Analysis (EDA)** on the data. Finally this data is passed through our information filtering and Content-based (Collaborative) algorithms which eventually provides a predictive strategy that is used as a **OTT Recommender system** for the user.

All this processed data is presented to the end user in the form of graphs and charts in our interactive dashboard.

1.2. Purpose

As OTT platforms are coming up with new ways to stand out among competitors by presenting original content, evidently more customers are being lost in deciding which platform would be suitable for their use. Moreover, most of the available recommendation systems are focused on suggesting the content but not the platforms that hold and provide those contents. To ease the choice dilemma, our web-app aims to present a guideline for choosing the appropriate OTT platform that fits one's personal preferences.

Features of our app not only allows the end users to compare different OTT platforms, but also explore each one of them individually. It would allow the users to be up to date with latest information about the OTTs. Thus, making their choice easier.

2. LITERATURE SURVEY

2.1. Existing problem

- **Scattered statistics and Content Recommendation:**
 - The statistics available on these OTT platforms are not aggregated and scatter all over the internet which makes comparison between these a tedious process. Moreover, most of the available recommendation systems are focused on suggesting the content but not the platforms that hold and provide those content.
- **Insignificant comparisons and lack of a comprehensive UI**
 - The available comparisons on the internet are mostly irrelevant to the actual user's needs and there is a void of a collective comprehensive user interface where the user might solve all their queries.

2.2. Proposed solution

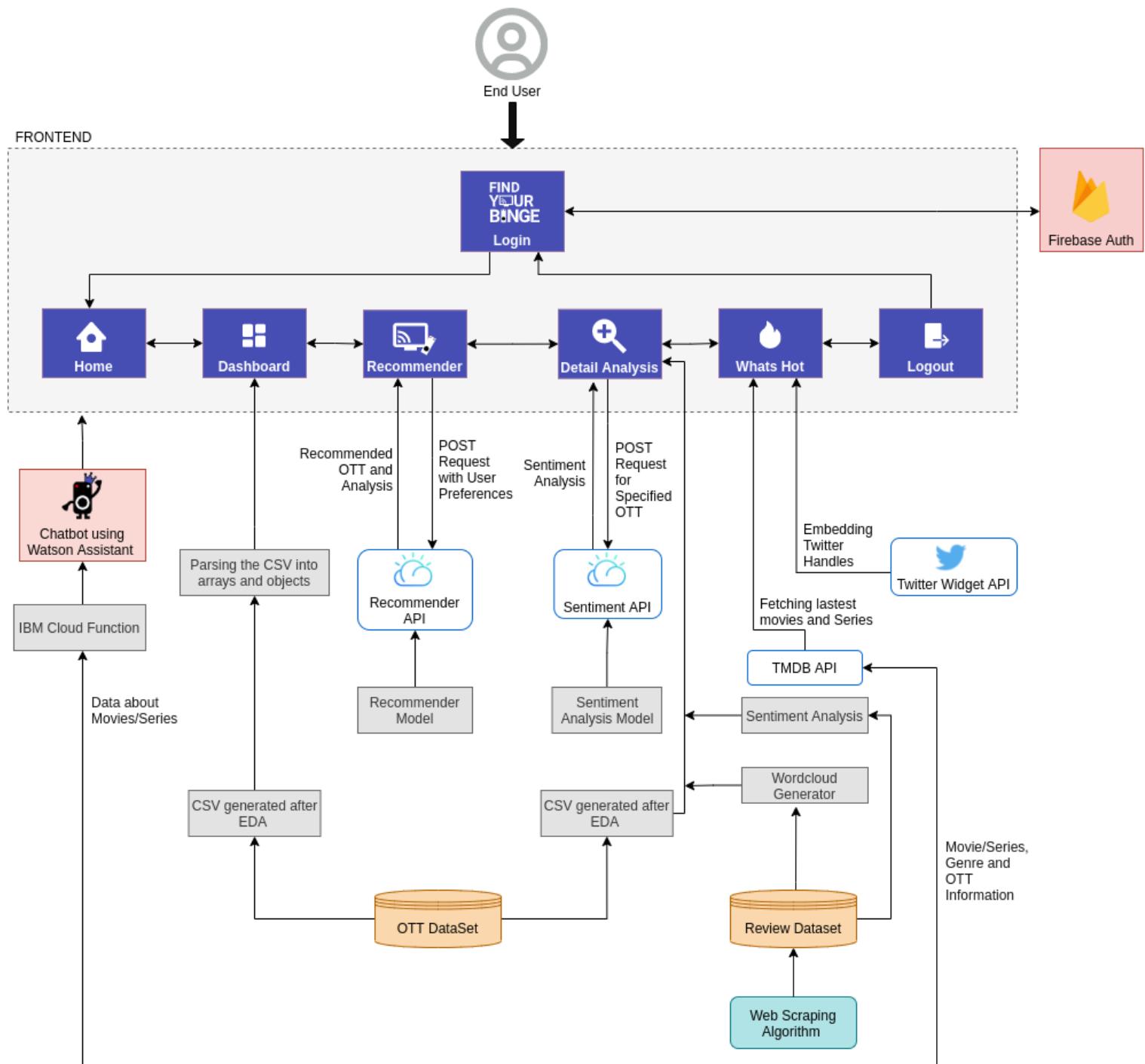
We propose a OTT analytics web app, **Find Your Binge** that can help bring upon a complete analysis of top OTTs, recommends the right OTT for the users, analyses what the public opinion is regarding all the OTTs in **real-time** and a lot more! Now the user can just sit back, and let FYB clarify all their doubts, the next time they sit to Binge watch!

Our solutions to the existing problems:

- **OTT Recommendation System:**
 - We have provided an OTT Recommender feature in our solution where the user can provide their preferences and the system will compute the best OTT for them. Moreover, the solution provides an aggregated comparison so that it is easy for the user to compare these platforms
- **Relevant Comparisons and Interactive User Interface:**
 - The solution focuses on comparing the OTTs on the most relevant parameters which are important while considering a suitable OTT platform such as, age rating, language, genre and IMDB Rating. These parameters are presented in an interactive manner using dynamically animated graphs and charts.

3. THEORETICAL ANALYSIS

3.1. Block diagram



*Note: Recommender API, Sentiment API and Web Scraping Algorithm is explained in the Flowcharts Section below

3.2. Software Designing

3.2.1. User Interface

- React

We have used React for our frontend Javascript Framework. We have used different npm modules along with React to increase the functionality of the web app. Following are the modules used:

- Create React App

Create React App or cra in-short is an officially supported way to create single-page React applications. It sets up all the required modules and configurations to get started with building a React application. This provided the base for the web app that we built.

- React Router Dom

React Router Dom is a library used to define multiple routes within a React application. We used this library to make the navigation of our web app. We made different routes for navigation to different pages and sections. This also provided us the functionalities to redirect to a different component and use URL paths.

- Node Sass

Sass is a preprocessing language that interprets itself into CSS. This provided the ability to write the styling of components in a much more logical manner.

- Highcharts

Highcharts is a charts and graphs library which provides interactive responsive charts. We used it to generate different types of graphs in our web app by providing different options to the graphs.

- GSAP

GSAP is a JS animation library. We use it to animate the authentication page transitions using the timeline feature.

- React Twitter Embed

This is a library used to embed twitter feeds and timelines into a react application. We used this in our What's Hot section to display tweets of the Top OTT Platform Handles.

- Redux

Redux is a state management tool that follows the FLUX pattern. It makes the state global and accessible from any component which avoids prop drilling.

The FLUX pattern utilises the unidirectional flow of data. Whenever an action (interaction from user) happens on the view (web app) the action is dispatched to a reducer (is a pure function which returns an updated state according to the action) and then to the store (global state) which is updated and reflects changes in the view.

- Redux Thunk

Redux Thunk is a library used to dispatch actions in an asynchronous manner. This helped us to make API calls in the actions and update the store when the JS promise was fulfilled.

- Reselect

Reselect is a library used to get values from the store in a convenient way. We made different selectors to select different slices of data from the store to use in different places.

- *Firebase Authentication*

We used Firebase (by Google) for the authentication part of our web app. It is an easy to integrate tool for different functionalities like OAuth, data storage and many more. We used the authentication with Email-Password and Google. Any auth process (login/sign up) on submission sends the data to the firebase server and returns an user object if authenticated or an error object if not authenticated. This object was used in the frontend to redirect to the home page on success or display of error message on failure.

- *API Usage*

We have used several custom-made and external APIs to introduce a variety of functionalities in our web app. The following are the APIs used:

- *Recommender API*

We have used a custom-made API for the recommender section of our app. This API takes input from the user. This input is the user's preferences about what they what in their OTT. After sending this input data to the API's server, this is processed and the results are returned to the frontend. Here we populate the returned data into tables and charts and display which is the best suited OTT for them according to out OTT Recommender System.

- *Real-time Tweet Analysis API*

We have used a custom-made API for getting Sentiment analysis of real-time tweets fetched at that moment of time. We call this API every 30 seconds to get a steady stream of real-time tweets' analysis to be showed in graphs.

- *The Movie DataBase (TMDB) API*

We have used this external API TMDB to retrieve any data related to latest trending movies and series and on which OTT platforms they are available. This is used in the What's Hot section and Chatbot of our web app.

- *CSV Usage*

After EDA (Exploratory Data Analysis), several CSVs were generated to be displayed in a visual manner in the web app. These CSVs are stored on the github repo and retrieved and parsed from there only. The parsed data is fed into the store using asynchronous actions.

- *Deployment on IBM Cloud*

This react application is deployed on **IBM Cloud** as a Cloud Foundry application.

3.2.2. Web Scraping and Word Cloud

- *Web Scraping*

We used an web automation tool, Puppeteer, to scrap the webpages (mentioned in the Appendix) for reviews of the OTTs. A simple script was written to go through a given number of paginated web pages and extract the reviews and title of reviews. Then these reviews were converted into a CSV File to be processed for sentiment analysis and word cloud generation.

- *Word Cloud Generation*

A word cloud is a collection, or cluster, of words depicted in different sizes which represents the frequency of each word.

We have created word clouds using the reviews of each OTT. These word clouds are interactive and user can hover over different words to see how often it's mentioned within the reviews of the specified OTT.

We have used NLTK(Natural Language Toolkit) library for cleaning, splitting strings into tokens and performing the frequency distribution to find the most used 100 words. The word clouds are generated with these words using the HighCharts Library.

3.2.3. Chatbot

- *IBM Watson Assistant*

We have added a chatbot to our web-app which is made using IBM Watson Assistant. It makes the interface more interactive and helps the user discover content on different OTTs.

We have used a combination of Dialog skill and Webhooks to build this Chatbot . The dialog skill interprets the user input, then directs the flow of the conversation. Webhooks is a mechanism that allows the chatbot to make a call out to IBM Cloud Web actions to resolve the user queries based on events in our dialog.

Our Chatbot has the following features :

- Providers information about different OTT Platforms
- User can query and find content on different OTT Platforms based on the genre of their choice
- If the user wants to watch a certain movie/series but has no idea on which platform it is available in India, our chatbot can find the provider for that content using just the title.
- Provides subscription packages of different streaming platforms.

- *IBM Cloud Functions*

We have used IBM Cloud Functions to create a web action. An action is a piece of code that performs one specific task. We have written this cloud function in Node.js which performs the task of calling out to External API (TMDB) to resolve user queries.

Actions can respond to events from IBM Cloud services and third party services using a trigger.

3.2.4. Natural Language Processing - NLP

We have created algorithms and functions which are used as a series of tools for further processing of natural language for cleaning and better understanding of the flow of data. The operations we have performed in this domain are -

- *Preprocessing*

We performed operations on tweets/reviews like-

- Deleting duplicated tweets/reviews, removing hashtags, retweets, links and punctuation characters
- Using lambda function to extract reviews based on important keywords, removing stopwords and finally converting to lowercase.
- The processed metadata is then sent to next step

- *Keyword extraction algorithm*

- It is used to transform a given text into a vector representation based on the frequency (count of occurrences) of each word that occurs in the entire text. It provides the capability to process your text data before generating the vector representation making it a highly flexible feature representation module for text.

- *Sentiment Analysis - VADER*

- **VADER** (Valence Aware Dictionary for Sentiment Reasoning): It is a model used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity of emotion. It uses a list of lexical features (e.g. words) which are labelled as positive or negative according to their semantic orientation to calculate the text sentiment. We chose this algorithm because it can be optimized for social media data and can yield fine results when used with unstructured data from Twitter.

- *Exporting the data*

- Final metadata sent to our algorithms for analysing tweets/reviews for classification in polarity, subjectivity, sentiment, negative, positive and neutral is sent to be displayed as WordCloud and consequent Graphs on the 'Sentiment Analysis' section of our site.
- Check out our [Sentiment Analysis- Watson Notebook](#)

3.2.4.1. OTT Review Analysis

After successfully scraping different opinions and reviews of people on the respective OTTs being analysed in this project, we cleaned and processed the data to perform sentiment analysis on the final corpus.

- *Sentiment Analysis*

- It refers to identifying as well as classifying the sentiments that are expressed in the text source. We have formulated our algorithm using **VADER**
- *WordCloud*
 - Furthermore, we calculated the *frequency of keywords* for building a WordCloud. Final metadata was sent to the React UI dashboard to be shown to the user as interpretable data.

3.2.4.2. LIVE @Twitter

Tweets are often useful in generating a vast amount of sentiment data upon analysis. This data is useful in understanding the opinion of the people about a variety of topics. Here, Live tweets are fetched from twitter using **Tweepy** which is an open source Python package that provides a convenient way to access the Twitter API with Python. After providing necessary credentials to authenticate our requests tweets are fetched from twitter. Pagination is used a lot in Twitter API development for example in iterating through timelines, user lists, direct messages, etc. In order to perform pagination and to make the process easier we are using **Tweepy's Cursor object**.

- *Flow of Twitter*
 - Request to fetch 100 tweets in real-time related to OTT-specific keywords
 - Cleaning the tweets, removing duplicates and applying preprocessing
 - *Sentiment Analysis*
 - The final Tweets are iterated through our software stack, which also involves **VADER**

3.2.5. OTT recommender:

- *Recommendation system:*
We have used content-based algorithms based recommender system which will process the user's preferences to help select an OTT platform that suits best to his/her preferences. In addition to the user's preferences we also input a impact factor which is similar to Latent matrix based technique wherein we produce the output according to the user given bias.
- *App creation:*
The Apps are created using Flask framework (by Python) which runs with Gunicorn Python WSGI HTTP server which handles POST requests The app provides an interface layer between the main python file and the React UI.

4. EXPERIMENTAL INVESTIGATIONS

The following were some observations made during the creation of this solution:

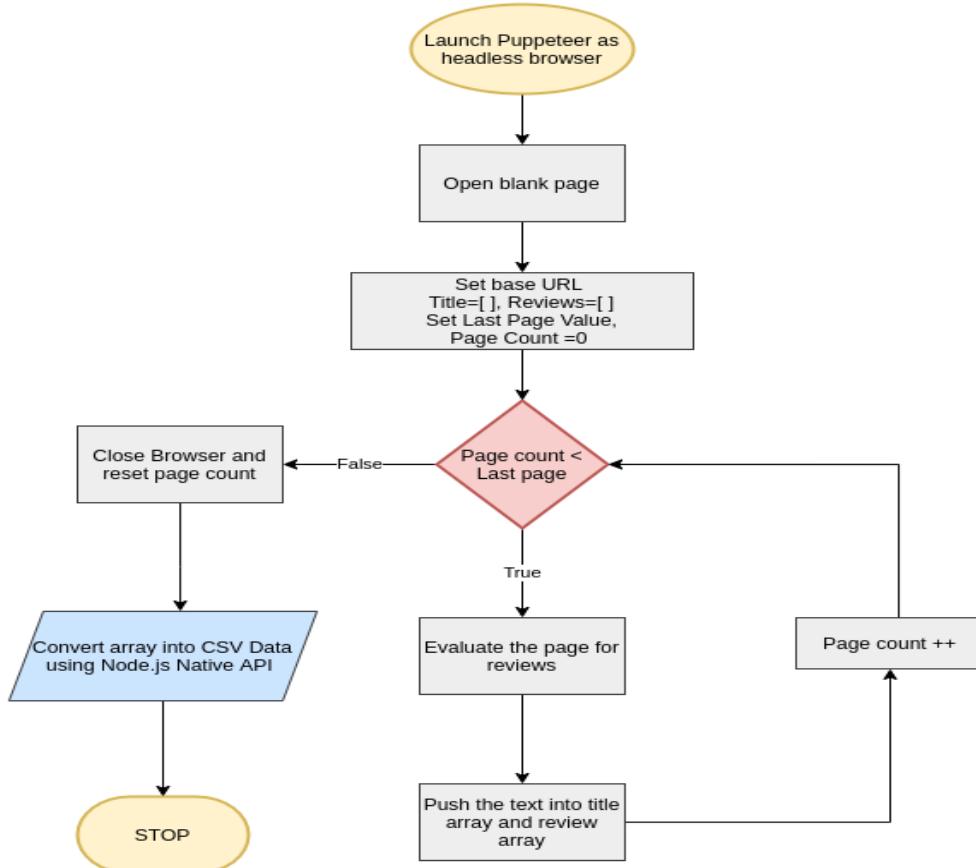
- According to the EDA performed on available dataset obtained **Netflix** has the highest

amount of content as compared to other OTT platforms.

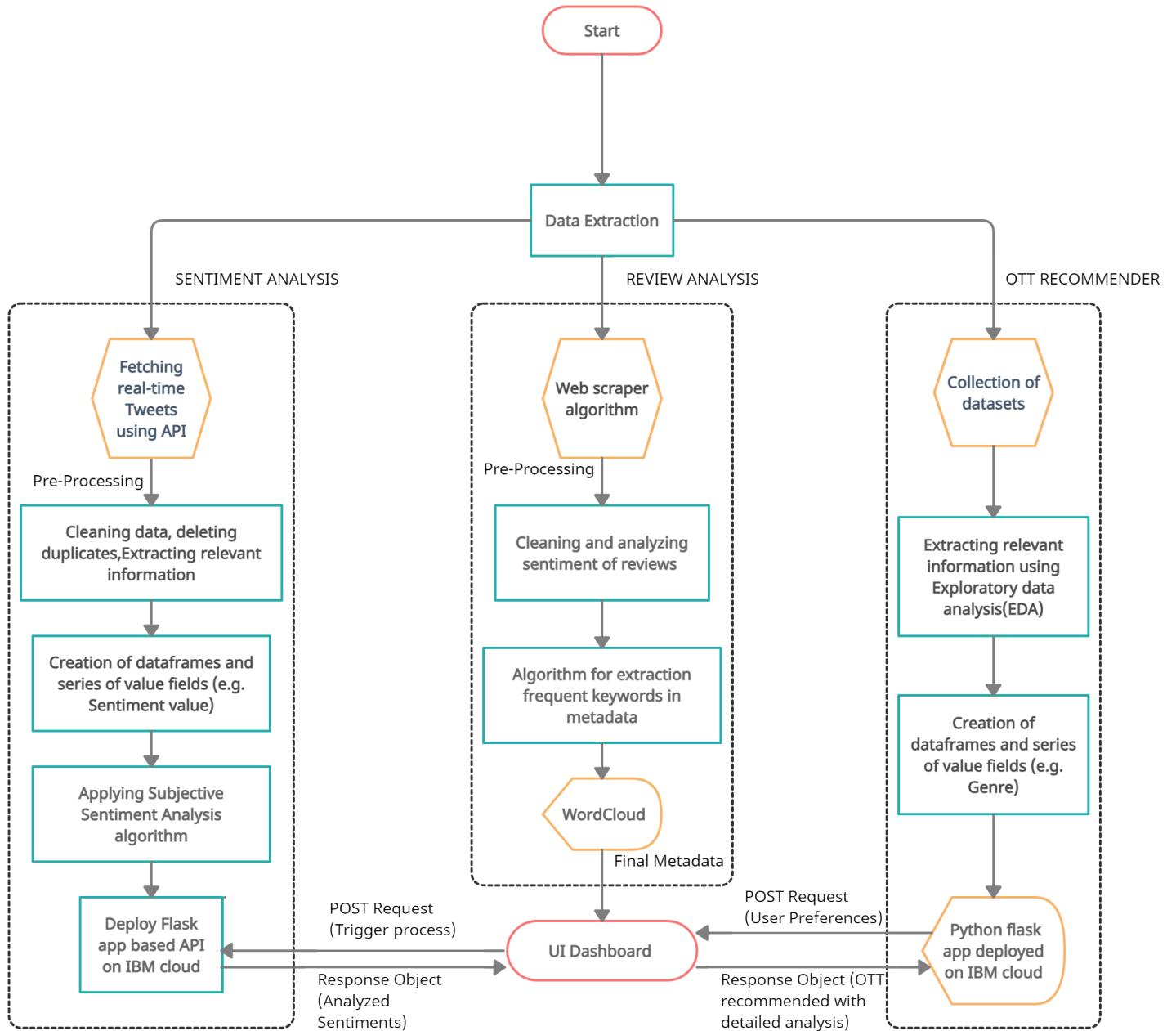
- The most popular genre that the OTTs include in their content library is **Drama** but predictably in Disney+ Hotstar platform **Family** shows and movies are prevalent as its audience comprises of younger age group.
- The most predominant language, unsurprisingly, is **English** but the second most varies on different platforms.
- Although Netflix has higher number of film titles than Disney+ Hotstar, still **Disney+ Hotstar** has higher **profit margin and subscriber count** in India. This may be due to lower pricing offered in Disney+. These values may change in the near future as Netflix is slowly pacing up in the viewership count.
- Covid-19 has been a catalyst in increasing adoption of OTT platforms due to which **subscriptions** of most of the OTT platforms have **increased** magnificently.
- Check out the [Preprocessing + EDA- Watson Notebook](#)

5. FLOWCHART

Web Scraper Algorithm:



Sentiment analysis, Review analysis and OTT recommender algorithm:



6. RESULT

The web app was successfully deployed on IBM Cloud as a Cloud Foundry Application and all deliverables are functional and LIVE.

Project Deliverables:

- **Homepage:** The Home page contains information about our web-app and all its features.

This page contains information about how OTT is better than Cable. It is an introductory page to our app.

- **The Dashboard:** The Dashboard contains different charts for visualising our comparative study between the top 3 OTTs based on the results of our EDA. The charts are interactive and you can remove or add the data with a click. The charts will help the user understand which OTT has what kind of content in a visual manner.
- **OTT Recommender:** The OTT Recommender uses Content-based algorithms based recommender system which will process the user's preferences to help select an OTT platform that suits best to users preferences.
- **Detail Analysis:** A Detailed Analysis of each OTT (Netflix, Prime Video and Disney+ Hotstar) is presented in a visual manner with extra details.
- **People's Opinion with LIVE Analysis:** Detail Analysis section has a real-time tweet analysis section and a section for the analysis of reviews posted by people about the respective OTTs.
- **What's Hot:** What's Hot provides the latest trending movies and series and information about them like genres and the OTT providers available. It also shows the tweets by the OTT platforms' twitter handles for the user to get up to date with the OTTs are uploading.
- **Watson Assistant Chatbot:** We have used the IBM Watson Assistant with IBM Cloud Functions to create a chatbot for the user to enquire on which platform a movie/series is available, movies/series according to genre and provider, information and subscription packages of an OTT.

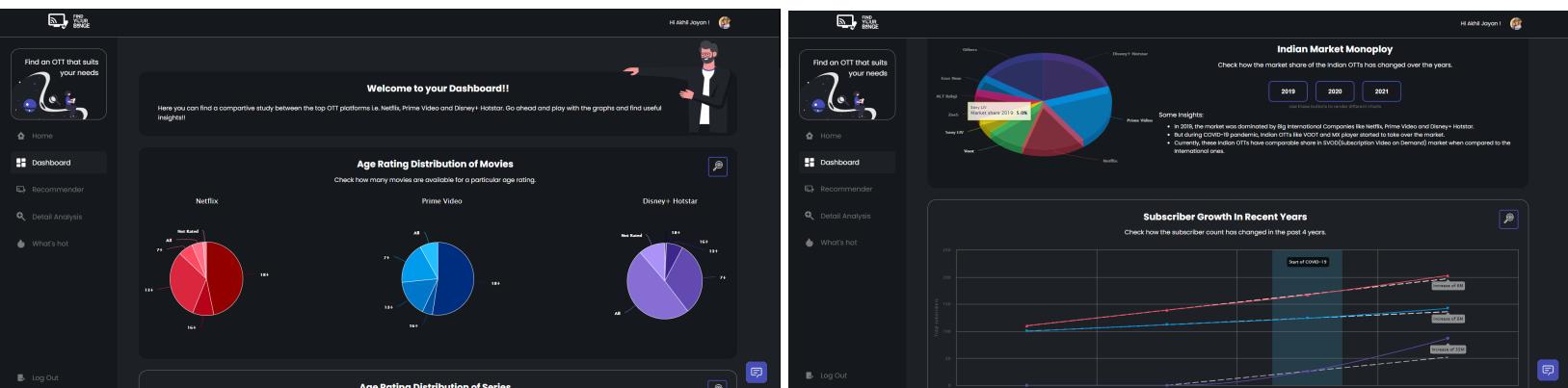
LIVE Web app Link : [Click Here](#)

Github Link : [Click Here](#)

Video Demo Link: [Click Here](#)

Some Screenshots:

For full dashboard screenshots: [Click Here](#)



Find on OTT that suits your needs

Let's Find Your Binge Together!!
Here you can find on OTT that best suits your preferences. You can take this simple quiz below to find the OTT of your dreams.

Select the options you would like in your OTT below..

What do you prefer to watch ? Set high priority: MOVIES, SERIES, BOTH
Content of which age rating would you prefer ? Set high priority: PG, 13+, 16+, UNIVERSAL

What genres do you like to watch? Set high priority: ACTION, ADVENTURE, ANIMATION, COMEDY, CRIME, DRAMA, DOCUMENTARY, FAMILY, FANTASY, HORROR, MYSTERY, ROMANCE, SCI-FI, SPORT, THRILLER
What languages do you watch ? Set high priority: ENGLISH, HINDI, TELUGU, GERMAN, FRENCH, ITALIAN, JAPANESE, KOREAN, MANDARIN, RUSSIAN, SPANISH

Content of what IMDB Rating do you like to watch ? Set High Priority: 8.0+, 7.0+, 6.0+, 5.0+

Test results are in!!
According to our algorithm, the best suited OTT platform for you is:

TAKE THE TEST AGAIN

In-Depth Series Analysis

Total number of Series
Genre Distribution

Parameter	NETFLIX	prime video	Disney+
FANTASY	6.0%	0.3%	7.2%
ANIMATION	8.9%	0.47%	11.20%

Comparison

Real-time Tweet Analysis
Check people's opinions analyzed by our Sentiment Algorithm.

Review Analysis
Customer Satisfaction Gauge

Wordcloud of reviews

Binge Master

Hii, I'm Binge Master. Can't decide what to watch and where to find it? Let me help you..

Here are the things that I can do for you..

- Information about different OTT Platforms
- Discover popular genre based content on top OTT Platforms
- Find where your favourite movies and TV Series are streaming
- Subscription Packages

Type something...

Built with IBM Watson®

Binge Master

What type of content are you looking for ?
Movies, TV Series

Select a genre to get TV Series recommendations based on your choices
Animation

Rick and Morty

Type something...

Built with IBM Watson®

Binge Master

The Lion King
Provider : Hotstar

Type something...

Built with IBM Watson®

7. ADVANTAGES & DISADVANTAGES

Advantages :

- Our application provides useful information to its users, by giving a comprehensive visual analysis of different OTT platforms.
- Our app offers an interactive chatbot which helps the user to discover new content.
- It also helps to alleviate the problem of information overload which is a prevalent phenomenon with information retrieval systems.
- Since users can understand the diversified content on OTTs, this brings great advantages as compared to conventional TV content.

Disadvantages:

- The recommender system is highly dependent on the dataset that the dataset might not contain data about some movies/series or data related to them.
- New and latest dataset could help in improving the comparative and detailed analysis.

8. APPLICATION

- Our web application helps our users to identify which OTT platform is best suited for them. It provides comparison between top OTT platforms along with an in-depth analysis of these OTTs. It also has a section where the users can get OTT Recommendations according to their preferences and a chatbot for their queries.
- It will be useful not only for the customers but the OTT companies as well, who can analyse their data to improve their content catering services .They can inspect the shortcomings in their content in comparison with other platforms.
- Our application promotes ease of use of OTTs, thereby increasing OTT business market.

9. CONCLUSION

As the COVID-19 situation seems to be easing a bit, people are still preferring to watch their favourite movies at the comfort of the home where they are safe. So the movie industry is also focusing on releasing more content sooner on OTTs. This will certainly make the consumer gravitate towards these online content catering services. Thus, our web app can be a beneficial tool to save a consumer's time and money by telling them which platform is the best for their needs. Moreover an OTT service can look over how it is competing in the market against its contenders.

10. FUTURE SCOPE

- **More data More accuracy**
 - With the advent of more and more dataset for our analytics team, we can improve upon all the aspects of our project. Currently, we are confined to the most popular OTTs as only their data is available on open source platforms. Therefore, if given the datasets of other OTTs we can expand upon our solution to recommend other OTTs as well.
- **Responsive Version**
 - In future, the web-app could be made responsive to phone screen sizes. This will allow users to use the app from the comfort of the phone.
- **Offline Support**
 - The app can be further evolved into a progressive web app(PWA). This will allow the users to use certain features of the app without a wifi connection.

11. BIBLIOGRAPHY

- Netflix Stats
 - <https://www.comparitech.com/tv-streaming/netflix-subscribers/>
 - <https://backlinko.com/netflix-users>
- Prime Video Stats
 - <https://backlinko.com/amazon-prime-users>
- Disney+ Hotstar Stats
 - <https://backlinko.com/disney-users>
- [OTT research](#)
- [VADER Research](#)

12. APPENDIX

A. Source Code

<https://github.com/smartinternz02/SBSPS-Challenge-5306-OTT-Platform-Analysis-Tool>

B. APIs Used

Recommender API

Real-Time Tweets' Sentiment Analysis API

[TMDB API](#)

C. Reviews Extracted From

[Netflix Reviews](#)

[Prime Video Reviews](#)

[Disney+ Hotstar Reviews](#)

D. Dataset Links

<https://drive.google.com/drive/folders/1ypjQBA7lbNBPQsgb4fHJIGzj5ItbbPag?usp=sharing>