



Analyzing the most preferred OTT platforms using ML

by

Ragunathan E S (2248014)
Aiswarya Antony (2248024)
Joan Jose M (2248038)

Under the guidance of
Dr. Ummesalma M

A Project report submitted in partial fulfillment of the requirements for
the award of the degree of Master of Science (Data Science) of CHRIST
(Deemed to be University)

April - 2023

TABLE OF CONTENTS

SL.No	Index	Page.No
-------	-------	---------

1	Introduction	3
2	Problem Statement	4
3	Aim and Objectives	5
4	Related works	6
5	Preliminaries	12
6	Data collection methodology and description of features	14
7	Proposed work	16
8	Results and Discussion	22
9	Conclusion and Future work	23
10	References	24

INTRODUCTION

The way we consume entertainment has drastically changed in recent years, and the rise of Over-the-top (OTT) platforms has played a significant role in this transformation. With the increasing number of OTT platforms available in the market, it becomes essential to analyze the most preferred ones to help users make informed decisions.

This project aims to use Machine Learning (ML) techniques to analyze the most preferred OTT platforms by considering various factors such as user ratings, genres, language, and content availability. By analyzing the data, we can identify the factors that drive user preferences and make recommendations for users who are looking for new OTT platforms to explore.

OTT, is a media service directly offered to viewers through the Internet, bypassing traditional cable, broadcast, and satellite TV providers who usually control or distribute such content. Unlike cable or satellite TV, OTT relies on high-speed internet to deliver film and television content. Cord cutters, who dislike paying for bundled content, often opt for OTT services like Netflix, Amazon Prime, MX Player, and Hotstar. However, OTT does not necessarily mean free, as it includes subscription-based services like Netflix and Amazon Prime. OTT allows viewers to selectively choose their providers at a much lower cost and avoid paying for bundles typical of cable and satellite TV. TVOD, or Transactional Video on Demand, refers to renting or purchasing movies through providers like Apple TV and YouTube.

The project involves data collection, preprocessing, feature engineering, model selection, and evaluation. Through this analysis, we hope to gain insights into the factors that influence user preferences and provide valuable information to both users and service providers.

PROBLEM STATEMENT

- Understand the Dataset & Features
- Perform Data Preprocessing Technique to Get Balanced Structured Data ● Perform Exploratory Data Analysis and Derive Valuable Insights
- Train and Test through Different Models for Better Prediction

- Compare the results from Support Vector Machine, Naive Bayes, Random Forest 4

AIM AND OBJECTIVES

Aim:-

The aim of this project is to use Machine Learning (ML) techniques to analyze the most preferred Over-the-top (OTT) platforms based on user ratings, content availability, genres, and language. By analyzing the data, we aim to identify the factors that drive user preferences and

provide recommendations for users who are looking for new OTT platforms to explore. Additionally, this project aims to provide valuable insights to OTT service providers on user preferences and areas of improvement to enhance user satisfaction.

Objectives:-

- To understand the reasons for the popularity of Online Video Streaming Services
- To analyze various factors that influence consumers to choose Online Video Streaming Services
- In classification, we are predicting the most preferred platform.

RELATED WORKS

1. The article "Data Analysis on OTT Platforms: Which Service Should I Choose? An Attempt to Resolve the Choice Dilemma of OTT Platforms through data analysis" explores the various OTT (over-the-top) platforms available in the market and helps readers make an informed decision about which one to choose. The author analyzes the subscription plans, content libraries, and user

ratings of popular platforms such as Netflix, Amazon Prime Video, and Hulu to determine which one offers the best value for money. The article also delves into the impact of user reviews on the success of these platforms and provides recommendations based on the data analysis. Overall, the paper provides a helpful guide for anyone looking to subscribe to an OTT platform.

2. The paper titled "Impact of Marketing Mix on Consumers Satisfaction - A Study of Consumers of OTT Platforms in India" examines the influence of marketing mix elements (product, price, place, and promotion) on consumer satisfaction in the OTT platform industry in India. The research is based on primary data collected through an online questionnaire from 350 OTT users in India. The data is analyzed using statistical techniques such as descriptive analysis, correlation analysis, and multiple regression analysis.

This study found that all four marketing mix elements (product, price, place, and promotion) significantly impact consumer satisfaction in the OTT platform industry. Product attributes such as content quality, variety, and originality significantly affect customer satisfaction among the four elements. The study also found that customers' price sensitivity plays a crucial role in their satisfaction with the OTT platform.

The findings of this study have practical implications for OTT platforms in India to improve their marketing mix strategies to enhance customer satisfaction.

The work recommends that OTT platforms focus on improving their content's quality, variety, and originality and offer flexible pricing options to attract and retain customers. Moreover, the study highlights the importance of effective promotion strategies to create brand awareness and attract new customers to the OTT platform.

6

3. The article "A cross-country analysis of the determinants of customer recommendation intentions for over-the-top (OTT) platforms" aims to identify the factors that influence customer recommendation intentions for Over-The-Top (OTT) platforms across different countries.

The authors surveyed 1,562 respondents from six countries and used a structural equation modeling approach to analyze the data. The study found that customer satisfaction and perceived value positively influence customer recommendation intentions for OTT platforms. In addition,

trust in the platform, social influence, and platform features were also found to impact customer recommendation intentions significantly.

However, the study also found that the effect of these factors on customer recommendation intentions varies across countries, indicating that cultural differences and market characteristics play an essential role in shaping customer behavior.

The authors conclude that OTT platforms must focus on delivering high-quality services that provide value to their customers and establish trust to enhance their customer recommendation intentions. They also suggest that platform features, such as ease of use and personalization, should be tailored to the specific needs and preferences of customers in different countries to maximize their impact on recommendation intentions.

4. The study "A Study on Customer Satisfaction on OTT Platforms during Covid-19 Pandemic Period" examines customers' satisfaction levels towards Over-The-Top (OTT) platforms during the COVID-19 pandemic. The study aims to understand the reasons behind the increase in the usage of OTT platforms during the pandemic and identify the factors that influence customer satisfaction.

The authors surveyed 350 respondents from different parts of India and analyzed the data using statistical methods. The study found that most respondents were satisfied with the content and services provided by OTT platforms during the pandemic. The study also identified several factors influencing customer satisfaction, including content quality, pricing, ease of use, and customer service.

This work concludes that OTT platforms have become an essential source of entertainment during the pandemic, and customer satisfaction is a critical factor in ensuring their long-term success. The study recommends that OTT platforms should focus on providing

7

high-quality content, transparent pricing, and excellent customer service to ensure customer satisfaction.

Overall, the study provides valuable insights into customers' satisfaction levels towards OTT platforms during the pandemic period and identifies the factors that influence their satisfaction. The study can be helpful for OTT platforms in formulating strategies to enhance

customer satisfaction and retention.

5. The article explores how Over-the-Top (OTT) platforms, such as Netflix and Amazon Prime, engage young consumers compared to traditional pay television services. The authors analyze the changing preferences of consumers and the use of gamification strategies by OTT platforms to attract and retain young audiences. The article suggests that gamification has become essential for engaging consumers in today's digital age. OTT platforms effectively use it to create immersive experiences and increase user engagement. The article also highlights the challenges traditional pay television services face in keeping up with these changing consumer preferences and calls for them to adopt similar gamification strategies to stay relevant.

6. The study "Viewer's Perception Towards 'OTT' Platform During Pandemic" investigates viewers' perception of Over-The-Top (OTT) platforms during the COVID-19 pandemic. The study aims to understand the factors that influenced viewers' decision to subscribe to OTT platforms during the pandemic and their level of satisfaction with the services provided.

The research used a survey method to collect data from 150 respondents from different parts of India. The study found that most respondents had subscribed to at least one OTT platform during the pandemic, primarily for entertainment and to pass the time. The study also identified several factors influencing viewers' decision to subscribe to OTT platforms, including content quality, pricing, ease of use, and availability of exclusive content.

This research further analyzed viewers' level of satisfaction with the services provided by OTT platforms during the pandemic. The study found that most respondents were satisfied with the content and services provided by the OTT platforms. However, the study also identified several areas where OTT platforms could improve, such as the availability of new and exclusive content, pricing, and customer service.

Overall, the work provides valuable insights into viewers' perceptions of OTT platforms during the pandemic. The study highlights the importance of content quality, pricing, and customer service in ensuring customer satisfaction and retention. The study can be helpful for OTT platforms in formulating strategies to enhance customer satisfaction and retention during and

after the pandemic.

7. The article "The Rise of OTT Platform: Changing consumer preferences" by Kaneenika Jain explores the growing popularity of OTT platforms and their impact on consumer behavior. The author explains that OTT platforms have gained momentum in recent years due to their affordability, convenience, and accessibility. With the rise of smartphones and high-speed internet, consumers increasingly turn to these platforms for entertainment and information.

The article highlights the shift in consumer preferences towards personalized and on-demand content made possible by OTT platforms. This has created new opportunities for content creators and distributors and challenges for traditional TV networks and cable providers. The author also discusses the impact of OTT platforms on advertising and the need for marketers to adapt to this changing landscape.

The article concludes with a discussion of the future of OTT platforms, including the potential for increased competition and consolidation in the market. Overall, the author emphasizes the importance of understanding consumers' changing preferences and behaviors to succeed in the OTT industry.

8. The article "Understanding platform strategies for consumer stickiness on OTT Platforms" explores the strategies OTT platforms employ to retain users.

In this article, the authors argue that stickiness is a critical factor for the success of OTT platforms. They identified six key strategies used by these platforms to enhance stickiness. These strategies are:

- Personalization: The platform personalizes the content for each user based on their viewing history, preferences, and demographics.

- Content diversification: The platform offers a wide variety of content to cater to the diverse tastes of its users.

- Pricing and packaging: The platform offers various subscription plans to suit different user segments' needs and budgets.

- User engagement: The platform engages users through social media, personalized recommendations, and user-generated content.
- Convenience: The platform provides easy access to content, with features such as download options, offline viewing, and cross-device synchronization.
- Technological innovations: The platform adopts the latest technologies, such as Artificial Intelligence, Virtual Reality, and Augmented Reality, to enhance the user experience. The authors also highlight the importance of data analytics and user feedback in improving the effectiveness of these strategies. Using these strategies, OTT platforms can create a sticky user experience that promotes user loyalty and maximizes revenue.

9. The paper "Is Over the Top Video Platform the Game Changer over Traditional TV Channels in India? A Niche Analysis" compares the popularity and consumption patterns of Over-The-Top (OTT) and traditional TV channels in India.

The authors surveyed 480 respondents and analyzed the data to identify the factors driving the growth of OTT platforms in India. They found that convenience, cost-effectiveness, and a more comprehensive range of content were the primary reasons users preferred OTT platforms over traditional TV channels. They also found that OTT platforms were more prevalent among younger audiences in urban areas.

The authors conclude that OTT platforms are a game-changer in the Indian media industry and can potentially disrupt the dominance of traditional TV channels. However, they also note that there are several challenges that OTT platforms need to overcome, including internet connectivity issues and the lack of regulation. Overall, the article provides valuable insights into the changing media consumption habits of Indian audiences and the implications for the Indian media industry.

10. The paper 'User Perception Towards OTT Video Streaming Platforms in Kerala (With Special Reference to Thrissur)' aims to understand users' perception towards Over-The-Top (OTT) video streaming platforms in Kerala, with a particular focus on the city of Thrissur. The

paper uses a questionnaire-based survey method to collect data from 200 respondents who are users of OTT platforms in Thrissur.

The work found that most respondents use OTT platforms for entertainment and offer a more comprehensive range of content and convenience than traditional television channels. The study also found that users are generally satisfied with the quality of content and the overall user experience of OTT platforms.

However, the study also highlights the concerns of users related to the cost of subscriptions, privacy, security issues, and the availability of restricted content. The study suggests that OTT platforms need to address these concerns to retain their user base and continue to grow in the Indian market.

Overall, the paper provides valuable insights into users' perception of OTT video streaming platforms in Kerala and can be helpful for industry practitioners, policymakers, and researchers interested in the OTT industry in India.

PRELIMINARIES

Our dataset contains categorical primary variables, so we are doing classification here. Some algorithms are good at dealing with categorical data.

We have used algorithms like Support Vector Machine, Naive Bayes, and Random Forest, which help to classify the most preferred OTT platforms.

1. SVM

Support Vector Machines (SVM) is a popular machine learning algorithm for classification and regression tasks. It is beneficial when dealing with complex datasets with many features or when the data could be more linearly separable. Here are some specific cases where SVM can be helpful:

Binary Classification: SVM is particularly effective where two data points must be separated.

High Dimensionality: SVM is particularly effective when dealing with datasets with many features. The kernel trick in SVM allows it to map the data into a higher-dimensional space, where it can be more easily separated.

Non-Linearly Separable Data: SVM is beneficial when dealing with data that is not linearly separable. The kernel trick used in SVM allows it to transform the data into a higher-dimensional space that can be linearly separable.

Outlier Detection: SVM can be used for outlier detection in datasets, where it can identify and remove the outliers from the dataset.

Text Classification: SVM is widely used in text classification tasks, such as sentiment analysis or spam detection. It can handle high-dimensional data and can be trained with relatively small datasets.

2. Naive Bayes

Naive Bayes is a simple yet powerful machine-learning algorithm commonly used in natural language processing, text classification, and spam filtering. It is based on Bayes' theorem and assumes that all features are independent. Naive Bayes works well when the following conditions are met:

- The input data is categorical or discrete.
- The dataset is large and diverse enough to train the model.
- The assumption of independence between features is close to reality.
- The problem is a classification problem where the output is a categorical variable. -

There are relatively few features compared to the dataset size.

Naive Bayes is particularly useful when the cost of misclassification is low and the training data is limited. It is a fast and efficient algorithm that can handle many features and is relatively insensitive to irrelevant features. However, its assumption of independence between components may not hold in some situations, leading to reduced accuracy.

3. Random forest

Random forest is a popular machine-learning algorithm for regression and classification problems. It is an ensemble learning method that combines multiple decision trees to improve the accuracy and robustness of the model. Random forest works well when the following conditions are met:

- The input data has categorical and numerical features.
 - The dataset is large and complex.
 - The problem is a classification or regression problem.
 - There are nonlinear relationships between the input features and the output variable. -
- The dataset contains missing values or noisy data.
- Interpretability of the model is a low priority.

Random forest is beneficial when dealing with high-dimensional datasets that contain a large number of features. It can handle missing values and noisy data without the need for preprocessing. Random forest is also robust to overfitting and is less sensitive to outliers. In addition, it can provide feature importance rankings, which can be helpful in feature selection and interpretation. However, the main disadvantage of random forest is that it can be computationally expensive and may require more time to train compared to other algorithms.

DATA COLLECTION METHODOLOGY AND DESCRIPTION OF FEATURES

For collecting the data from OTT platform users, the following steps were performed:

1. Preparation of Questionnaire
2. Google Form Creation
3. Circulating Google Forms

A questionnaire was created, and responses about User Satisfaction Feedback on OTT platforms using Google Forms were collected.

Female	Age	Occupation	How many children	Which 2nd job	Do you	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which 2nd job	How often	What	Which
--------	-----	------------	-------------------	---------------	--------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	---------------	-----------	------	-------

Fig: User Satisfaction Feedback responses were collected using <https://forms.gle/qxiLe1ZTGSV7OVfXA>

The response data contained 243 instances and 31 attributes.

The attribute includes the following features:

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 243 entries, 0 to 242
Data columns (total 31 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Gender                                     243 non-null    object
1   Age                                       243 non-null    object
2   Occupation                               243 non-null    object
3   No. of platforms using                   243 non-null    object
4   Preferred platform                       243 non-null    object
5   Preferred language                       243 non-null    object
6   Viewing habits influence                 243 non-null    object
7   Avg. time spent                         243 non-null    object
8   Often usage                             243 non-null    object
9   Preference                               243 non-null    object
10  Devices used                             243 non-null    object
11  Cancelled subscription due to high cost  243 non-null    int64
12  Shared Account info                     243 non-null    int64
13  Dissatisfied with availability of content 243 non-null    int64
14  Plan preference                           243 non-null    object
15  Subscribed to watch specific content     243 non-null    int64
16  Cancelled subscription due to streaming quality 243 non-null    int64
17  Factors considered                       243 non-null    object
18  Free trial                               243 non-null    int64
19  Reasonable price                         243 non-null    int64
20  Switched platform                       243 non-null    int64
21  Content not available in your country     243 non-null    int64
22  VPN access for content unavailable        243 non-null    int64
23  Rating & reviews for decision            243 non-null    int64
24  Buffering while streaming                 243 non-null    object
25  Satisfied with availability of subtitles & closed captions 243 non-null    int64
26  Before Covid                             243 non-null    int64
27  Increase usage during lock down          243 non-null    object
28  OTT instead of Cinema Hall               243 non-null    object
29  Cheaper than traditional media           243 non-null    int64
30  Frequent theatre visit as before         243 non-null    object
dtypes: int64(14), object(17)
memory usage: 59.0+ KB

```

15

PROPOSED WORK

Since we collected the data via Google form, we need to clean the data before building the model.

I. Preprocessing

It is crucial in machine learning because the data quality used to train a model can significantly impact its performance. Here are the main steps in data preprocessing.

Steps involved in Preprocessing

a) Importing the data:

We must import the data into Python using a library like Pandas. Pandas is a popular library for working with data frames in Python.

243 × 31

b) Removing unnecessary columns or rows:

We may have some columns or rows in our dataset that are irrelevant to our classification task. So, we can remove those using the **drop()** method in Pandas. Since here, we are not dropping any columns or rows because everything is considered.

16

c) Handling missing data:

If there are missing values in the dataset, we need to handle them before training our model. We can use techniques like mean imputation, median imputation, or forward/backward filling to replace missing values.

Here, our dataset contains categorical primary variables, so if there is any missing data, we can go with mode imputation to replace missing values.

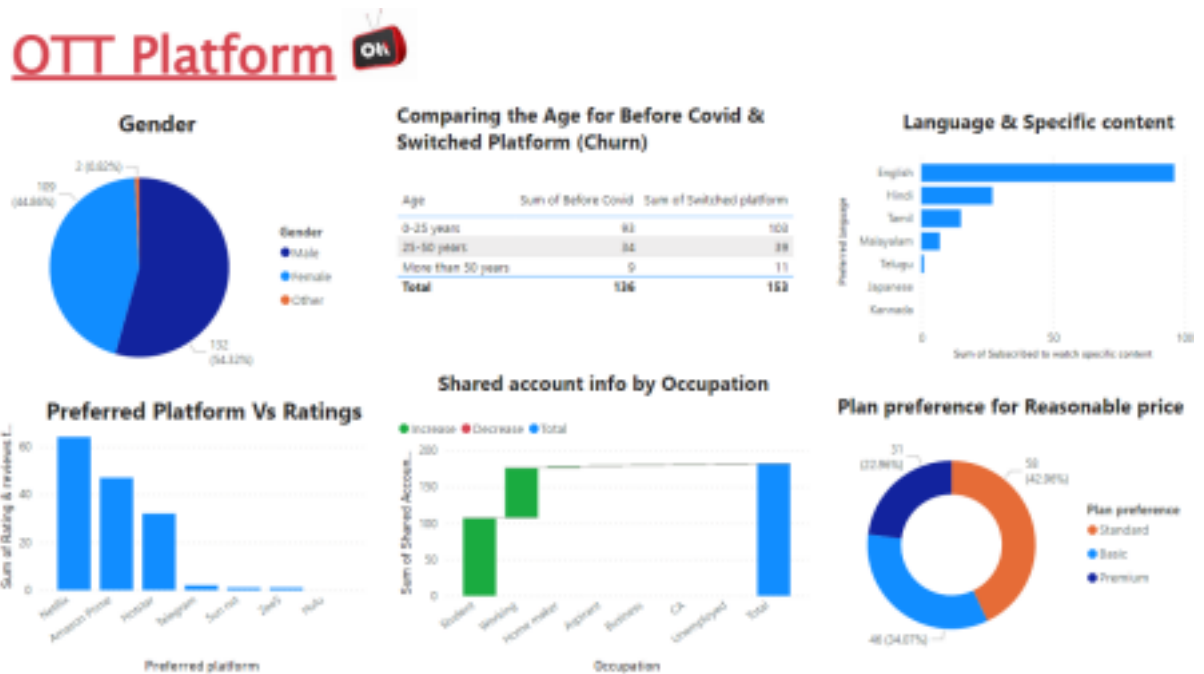
Treating the missing values using SimpleImputer using 'most_frequent' is nothing but mode.

d) Understanding the data:

Here, we can use Exploratory data analysis (EDA) to understand the data better. We'll use the Correlation heatmap to find linear and non-linear relationships between variables. And further,

we have done a bar chart, pie chart, And distribution plot to determine whether the data is normally distributed or skewed.

We have also used tools like Power bi to understand our data easily.



e) Identifying the Outliers:

We need to proceed by checking whether our data has outliers or not. The best way to detect outliers includes Z-score, box plot, Tukey's method, etc.,

17

f) Converting categorical variables to numerical variables:

If we have a categorical variable in our dataset, we must convert them to numerical variables so our model can process them. Many techniques include One-hot encoding, Label encoding, Ordinal encoding, Target encoding, and Binary encoding. These techniques help us convert categorical to numerical, i.e., binary data (0's & 1's).

g) Standardizing or normalizing the data:

Our dataset may have features with significantly different scales. We can use techniques like standardization or normalization to ensure all elements are on a similar scale.

h) Selecting Feature and target variable:

Features (X) - input variable or attribute used to describe the data. They are the characteristics or properties of the data that machine learning uses to make predictions or classifications. Target (y)

- output variable that the machine learning algorithm is trying to predict or classify. X- except

‘Preferred Platform.’

y- ‘Preferred Platform.’

i) Splitting the data into Training and testing datasets:

Finally, we must divide our data into a training dataset (70%) and a testing dataset (30%). The training dataset will be used to train our model, while the testing dataset will be used to evaluate its performance.

II. Model Building and Evaluation

Model building involves selecting the appropriate machine learning algorithm, training it on a labeled dataset, and fine-tuning it to optimize performance. The main goal of model building is to develop a predictive model that can generalize well on unseen data. This requires careful selection of features (input variables), data cleaning, preprocessing, and model selection. Model building is an iterative process requiring continuous improvement and refinement to achieve better results.

18

Supervised learning:-

Classification is a common problem in supervised learning where the goal is to predict a categorical label or class for a given input data point. The labeled dataset contains input data (also known as features) and corresponding output labels (also known as target variables) which the algorithm uses to learn a mapping function between the input and output data.

Classification Report of various Algorithm

1. SUPPORT VECTOR MACHINE:

Here we got an accuracy score of 55% for SVM and while doing hyperparameter tuning in SVM

we got an accuracy of 65.3% which is more than SVM.

	precision	recall	f1-score	support
Amazon Prime	0.58	0.65	0.61	23
Hotstar	0.42	0.48	0.44	21
Hulu	0.00	0.00	0.00	1
Netflix	0.65	0.54	0.59	28
accuracy			0.55	73
macro avg	0.41	0.42	0.41	73
weighted avg	0.55	0.55	0.55	73

Fig. : SVM Classification Report

2. NAIVE BAYES:

i) Gaussian:

It works by first calculating the mean and variance of each feature for each class in the training set and we got an accuracy score of 49% for gaussian.

	precision	recall	f1-score	support
Amazon Prime	0.43	0.13	0.20	23
Hotstar	0.70	0.33	0.45	21
Hulu	0.00	0.00	0.00	1
Netflix	0.46	0.93	0.62	28
accuracy			0.49	73
macro avg	0.40	0.35	0.32	73
weighted avg	0.51	0.49	0.43	73

Fig. : Gaussian Naive Bayes Classification Report

19

ii) Bernoulli:

It works by first calculating the probability of each feature being 1 or 0 for each class in the training set. However, it may not be suitable for datasets with continuous input features, we got an accuracy score of 62%.

	precision	recall	f1-score	support
Amazon Prime	0.64	0.78	0.71	23
Hotstar	0.50	0.38	0.43	21
Hulu	0.00	0.00	0.00	1
Netflix	0.66	0.68	0.67	28
accuracy			0.62	73
macro avg	0.45	0.46	0.45	73
weighted avg	0.60	0.62	0.60	73

Fig. : Bernoulli Naive Bayes Classification Report

iii) Categorical:

While we are doing with categorical naive bayes we got the highest accuracy score 73% because it assumes categorical input features that follow a categorical distribution.

	precision	recall	f1-score	support
Amazon Prime	0.78	0.78	0.78	23
Hotstar	0.68	0.62	0.65	21
Hulu	1.00	1.00	1.00	1
Netflix	0.70	0.75	0.72	28
accuracy			0.73	73
macro avg	0.79	0.79	0.79	73
weighted avg	0.73	0.73	0.73	73

Fig. : Categorical Naive Bayes Classification Report

iv) Complement:

It is similar to the Multinomial Naive Bayes algorithm but differs in the way it calculates probabilities. For complement we got an accuracy score of 63%.

Fig. : Complement Naive Bayes Classification Report

v) Multinomial:

It is particularly useful when dealing with high-dimensional datasets with discrete features, such as text data. For multinomial we got an accuracy score of 62%.

	precision	recall	f1-score	support
Amazon Prime	0.71	0.74	0.72	23
Hotstar	0.62	0.38	0.47	21
Hulu	0.00	0.00	0.00	1
Netflix	0.61	0.71	0.66	28
Telegram	0.00	0.00	0.00	0
accuracy			0.62	73
macro avg	0.39	0.37	0.37	73
weighted avg	0.63	0.62	0.61	73

Fig. : Multinomial Naive Bayes Classification Report

3. RANDOM FOREST:

However, it may not perform well on small datasets, as the algorithm requires a large number of decision trees to achieve good performance. This also might be a reason for getting an accuracy score of 52%.

	precision	recall	f1-score	support
Amazon Prime	0.57	0.70	0.63	23
Hotstar	0.00	0.00	0.00	21
Hulu	0.00	0.00	0.00	1
Netflix	0.50	0.79	0.61	28
accuracy			0.52	73
macro avg	0.27	0.37	0.31	73
weighted avg	0.37	0.52	0.43	73

Fig. : Random Forest Classification Report

RESULTS AND DISCUSSION

While comparing accuracy scores of all the three models i.e. Support Vector Machine, Naive Bayes and Random Forest, we found that among all the three models:

Model Used	Accuracy Score
Support vector Machine	52%
Categorical Naive Bayes	73%

Random Forest	55%
---------------	-----

CONCLUSION AND FUTURE WORK

In conclusion, the analysis of the most preferred OTT platforms using Machine Learning has shown that Netflix is the most popular platform among users followed by Amazon Prime. The analysis was based on various factors such as user reviews, content availability, and subscription pricing. The machine learning algorithms used for the analysis were Support Vector Machine, Naive Bayes and Random Forests, which proved to be effective in predicting the popularity of the OTT platforms.

The analysis also revealed that the availability of exclusive content and a user-friendly interface were crucial factors in determining the popularity of OTT platforms. Additionally, the study showed that the popularity of OTT platforms varied based on different geographical locations. Overall, the use of machine learning in analyzing the most preferred OTT platforms provides valuable insights that can help OTT providers in enhancing their services to meet the expectations of their users.

- Most of the viewers of OTT platforms were students, this might affect their academic performance and study pattern.
- It is a huge concern that students deviate from studies and spend more hours on OTT platforms which in return affects their grades.
- Since most of the preferred platforms are Netflix and Amazon Prime, in future we can include steps to improvise other platforms than the above ones.

REFERENCES

- [1] Kiran K, Ajay Kumar M, “A study on customer’s perception on OTT platforms in recent times”, *International Journal of Research Publication and Reviews*, vol 3, pp 1264-1266, April 2022 [2] Kamle Uday Kumar (2020), A consumer preference towards OTT platform during pandemic, *studocu* [3] Rick Kim (2021, May 20), Data Analysis on OTT Platforms: Which Service should I choose?, *Towards Data Science*
- [4] Chyoug Hwa Chen, I Fei Chen, Ruey Chyn Tsaur, Li Yun Chui, User behaviors analysis on OTT platform with an integration of technology acceptance model, *Springer Link*, February 2023 [5] Shilpa Parihar, Kamal Kant, Impact of Marketing Mix on Consumers Satisfaction - A Study of Consumers of OTT Platforms in India, *ResearchGate*, August 2021
- [6] Gurpreet Kour, Bhavika Chhabria, Understanding platform strategies for consumer stickiness on OTT platforms, *Journal of Indian Business Research*, vol 14, November 2022
- [7] Kaneenika Jain, The rise of OTT platform: Changing Consumer preferences, *EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal*, vol 7, June 2021 [8] Sujith T S, M Sumathy, User Perception towards OTT video streaming platforms in Kerala (with special reference to Thrissur), *Open Access Journal*, vol 2, August 2021
- [9] S Gomathi, N Vijaitha Christy, Viewer’s perception towards OTT platform during pandemic (with special reference to Coimbatore), *International Journal of Creative Research Thoughts (IJCRT)*, vol 9, August 2021

