International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

ENTERTAINMENT OTT CONTENT ANALYTICS AND PREDICTION

[1] R Ramu, [2] G Umesh Chandra, [3] K Bheemesh, [4] B Sony
[1]AssistantProfessor, [234]Students
Department of Data Science
Siddhartha Institute of Technology & Sciences, Narapally
RAMURAMAVATH@siddhartha.org.in, 23TQ1A6757@siddhartha.co.in,
23TQ1A6731@siddhartha.co.in, 23TQ1A6711@siddhartha.co.in,

ABSTRACT

The rapid growth of Over-The-Top (OTT) platforms has resulted in a large amount of content and user interaction data. OTT platforms provide movies, web series, documentaries, and other digital content across different languages, genres, and regions. Analyzing this data can help understand content popularity, audience preferences, viewing patterns, and platform performance.

The proposed Entertainment OTT Content Analytics and Prediction system is designed to analyze OTT content and generate meaningful insights from historical data. The system processes attributes such as content title, genre, language, release year, duration, content type, ratings, number of views, popularity, and audience engagement. Data preprocessing techniques are applied to clean and organize the collected information.

The system performs descriptive and exploratory analysis to identify popular genres, languages, content types, release trends, audience ratings, and engagement patterns. Interactive dashboards provide visual representations using charts, graphs, tables, filters, and KPI cards. Users can compare content based on ratings, popularity, release year, genre, language, and other available attributes.

The prediction module uses historical OTT data and Machine Learning techniques to estimate content popularity, ratings, or audience engagement. Predictive models can identify relationships between content characteristics and audience response. The system can also incorporate sentiment analysis of user reviews to understand positive, neutral, and negative audience opinions.

Overall, the proposed system combines OTT content analytics, audience behavior analysis, visualization, and predictive modeling in a unified platform. It can help content analysts and OTT businesses understand historical performance and audience preferences. Future enhancements can include personalized recommendations, real-time viewing analytics, churn prediction, content demand forecasting, and advanced audience segmentation.

I. INTRODUCTION

OTT platforms have transformed the entertainment industry by providing users with on-demand access to movies, web series, documentaries, and other digital content. The increasing adoption of streaming services generates large volumes of information related to content, ratings, viewing behavior, subscriptions, and audience engagement. This data provides opportunities for detailed analysis of digital entertainment trends.

Content performance depends on several factors such as genre, language, release year, duration, content type, ratings, marketing, and audience preferences. Understanding these factors can help OTT platforms analyze which types of content receive greater audience engagement. Historical content data can also reveal changes in viewing preferences over time.

Traditional OTT platforms mainly focus on content discovery, streaming, user accounts, subscriptions, and recommendations. Although these systems provide basic

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

content information and personalized suggestions, detailed analytical information may not be directly available to users or business analysts. Separate analytics tools may therefore be required for deeper content analysis.

Data analytics can be used to calculate important OTT performance indicators such as average ratings, total views, popularity scores, content distribution, and engagement levels. Visualization techniques can represent these indicators through dashboards, charts, graphs, and tables. This makes large and complex OTT datasets easier to understand.

Machine Learning can further enhance OTT analytics by predicting content popularity, ratings, audience engagement, or demand. Historical content characteristics can be used to train predictive models. The proposed system integrates data preprocessing, content analysis, audience analytics, visualization, and Machine Learning-based prediction to provide a comprehensive OTT analytics framework.

II. LITERATURE SURVEY

OTT content analytics has become an important research area because streaming platforms generate large volumes of content and audience-related data. OTT datasets may contain information such as content title, genre, language, release year, duration, ratings, views, and popularity. Researchers use these attributes to identify content trends and understand audience preferences.

Content popularity analysis is commonly used to evaluate the performance of movies and web series. Metrics such as views, ratings, watch time, popularity scores, and engagement can be analyzed to determine content performance. Comparing these metrics across different content categories can reveal patterns in audience interest.

Genre and language analysis are also important areas of OTT research. Audiences may have different preferences for genres such as action, comedy, drama, thriller, romance, animation, and documentary. Language-based analysis can identify regional preferences and changes in content consumption across different language groups.

Audience reviews and feedback provide additional information about content quality and viewer satisfaction. Natural Language Processing techniques can be applied to textual reviews to classify sentiment as positive, negative, or neutral. Keyword and topic analysis can also identify common opinions related to acting, storyline, direction, music, and overall content quality.

Recommendation systems are widely studied in the OTT domain. Content-based filtering, collaborative filtering, and hybrid recommendation techniques can recommend movies or series based on user preferences and historical viewing behavior. Machine Learning can also be used to predict ratings, content popularity, and potential audience engagement.

Modern OTT analytics systems combine descriptive analytics, visualization, recommendation systems, sentiment analysis, and predictive modeling. Interactive dashboards can provide insights into content performance, genre trends, audience ratings, and engagement. The proposed framework follows this approach by integrating OTT content analytics with audience analysis and prediction capabilities.

III. SYSTEM ANALYSIS

The Entertainment OTT Content Analytics and Prediction system is designed to process OTT content and audience data and generate meaningful insights. The system accepts information such as content title, genre, language, release year, duration, content type, ratings, views, and audience engagement. The input dataset is validated to ensure that required fields and supported formats are available. A preprocessing

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

module handles missing values, duplicate records, invalid entries, and inconsistent data formats. Date and categorical attributes are standardized to support reliable analysis. Feature engineering creates useful attributes such as release period, popularity category, rating category, and engagement indicators. The analytics engine calculates content counts, average ratings, views, popularity, and audience engagement metrics. Genre and language analysis identifies popular content categories and regional preferences. Time-based analysis examines changes in OTT content and audience preferences across different years. Visualization components present results through charts, graphs, tables, and KPI cards. A Machine Learning module can be trained to predict content popularity, ratings, or audience engagement. The prediction module generates estimated outcomes for new content based on selected features. The final dashboard presents analytical and predictive results through an interactive interface.

Existing System

Existing OTT platforms provide users with access to movies, web series, documentaries, and other digital content. These platforms generally include content browsing, search, ratings, reviews, subscriptions, and recommendation features. Although these features support content discovery, detailed analytical information may be limited. Users and analysts may not have direct access to comprehensive content-performance comparisons. Comparing genres, languages, ratings, and release trends can require separate data-analysis tools. Audience engagement patterns may not be presented in a detailed analytical dashboard. Reviews may be available but may not always be systematically analyzed using sentiment techniques. Historical changes in content popularity can also be difficult to examine through standard interfaces. Traditional recommendation systems may provide suggestions without exposing detailed analytical information. Predictive analysis of content performance may require separate Machine Learning applications. Data from different OTT sources may also have different formats and attributes. These limitations create a need for a unified system that combines content analytics, audience analysis, visualization, and predictive capabilities.

Disadvantages of Existing System

- Limited detailed OTT content analytics.
- Difficult to analyze large content datasets manually.
- Limited content-performance comparison.
- Limited genre-wise and language-wise analysis.
- Historical trends may be difficult to identify.
- Limited audience engagement analysis.
- Reviews may remain largely unstructured.
- Limited sentiment analysis.
- Limited predictive content analysis.
- Separate tools may be required for advanced analytics.
- Limited customizable analytical dashboards.
- Difficult to analyze multiple OTT content attributes together.

Proposed System

The proposed Entertainment OTT Content Analytics and Prediction system provides a centralized platform for analyzing OTT content and audience-related data. Users can upload historical OTT datasets or connect the system to an authorized data source. The data validation module checks the structure, required fields, data types, and basic

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

data quality. The preprocessing module removes duplicate records and handles missing, invalid, and inconsistent information. Feature engineering creates useful attributes such as content type, release period, rating category, popularity level, and engagement indicators. The analytics module calculates important metrics including content count, average rating, views, popularity, and audience engagement. Genre and language analysis identifies popular content categories and regional preferences. Users can compare movies and web series based on ratings, popularity, release year, language, and genre. Time-series analysis identifies changes in content availability and audience preferences over time. Natural Language Processing can be incorporated to analyze audience reviews and determine sentiment patterns. Machine Learning models can predict content ratings, popularity, or engagement using historical data. Interactive dashboards present analytical results through KPI cards, charts, graphs, tables, and filters. The proposed system therefore combines OTT analytics and intelligent prediction in a unified platform.

Advantages of Proposed System

- Provides centralized OTT content analysis.
- Automates data preprocessing.
- Supports genre-wise analysis.
- Supports language-wise content analysis.
- Provides content-performance comparison.
- Identifies audience preference patterns.
- Analyzes ratings and popularity.
- Supports time-based trend analysis.
- Can incorporate audience sentiment analysis.
- Provides interactive dashboards.
- Supports predictive content analysis.
- Provides a foundation for personalized recommendations.

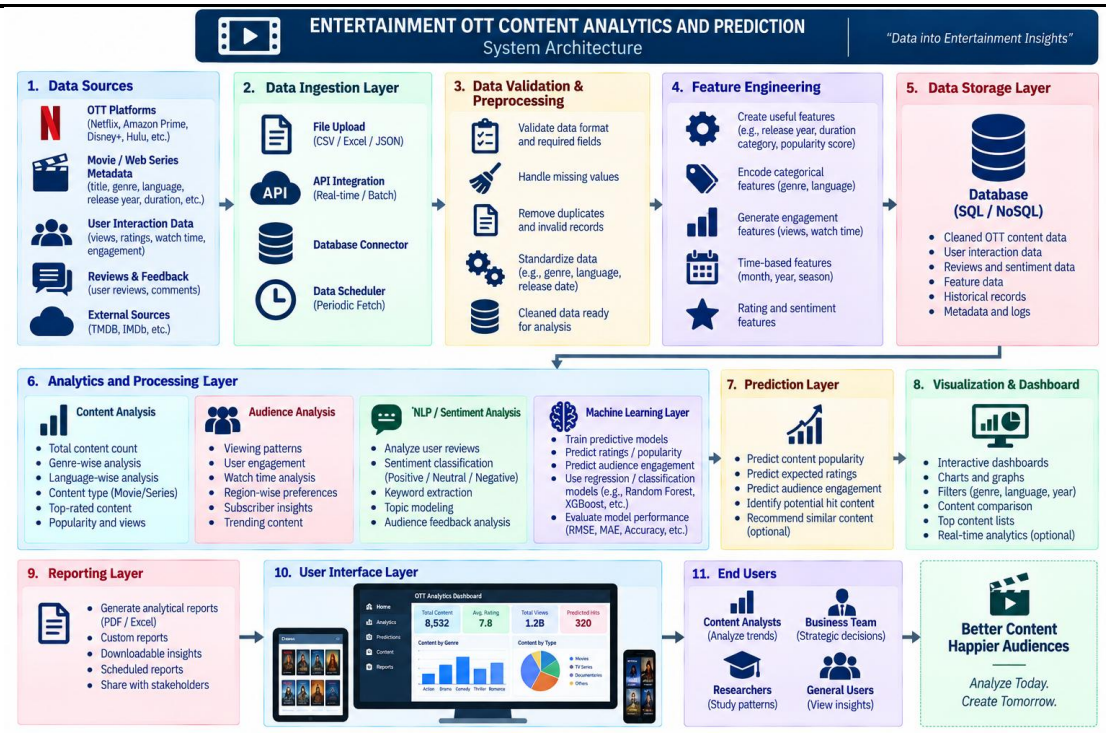
IV. METHODOLOGY

The methodology begins with collecting OTT content and audience data from authorized datasets or data sources. The collected data is validated to check its structure, required attributes, data types, and completeness. Data preprocessing is performed to remove duplicate records and handle missing, invalid, and inconsistent values. Numerical attributes such as ratings, views, duration, and popularity scores are standardized when required. Categorical attributes such as genre, language, and content type are transformed into suitable formats for analysis. Feature engineering creates useful attributes such as release period, rating category, popularity level, and engagement indicators. Exploratory Data Analysis is conducted to understand content distribution and audience preferences. Genre, language, and content-type analysis is performed to identify popular content categories. Time-based analysis identifies changes in content availability, ratings, and audience engagement. Natural Language Processing can be applied to reviews to perform sentiment and keyword analysis. Machine Learning models can be trained using historical content features to predict ratings, popularity, or audience engagement. Suitable evaluation metrics are used to assess the predictive model. Finally, the analytical and predictive results are displayed through an interactive OTT analytics dashboard.

SYSTEM ARCHITECTURE

The system architecture consists of several interconnected layers that process OTT content and audience data and generate analytical and predictive insights.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



The Data Source Layer provides movie, web-series, rating, review, view, genre, language, and engagement information from authorized sources. The Data Input Layer accepts information through CSV files, Excel files, databases, or APIs. The Data Validation Layer checks file formats, required attributes, data types, and basic data quality. The Data Preprocessing Layer handles missing values, duplicate records, invalid entries, and inconsistent formats. The Feature Engineering Layer creates attributes such as content type, release period, rating category, popularity level, and engagement indicators. The Data Storage Layer securely stores cleaned OTT records and historical information. The Analytics Processing Layer calculates content counts, ratings, views, popularity, and engagement metrics. The Content and Audience Analysis Layer compares genres, languages, content types, audience ratings, and engagement patterns. The NLP Layer analyzes audience reviews and identifies sentiment, keywords, and common feedback topics. The Machine Learning Layer trains and evaluates models for rating, popularity, or audience-engagement prediction. The Prediction Layer generates estimated content performance for new records. The Visualization Layer presents results through interactive dashboards, charts, graphs, tables, and KPI cards. The Reporting Layer generates analytical summaries and reports for authorized users.

V. RESULT AND OUTPUT

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

VI. CONCLUSION

The Entertainment OTT Content Analytics and Prediction system provides a data-driven platform for analyzing digital entertainment content and audience preferences. The system transforms raw OTT data into meaningful insights through data preprocessing, analytics, visualization, and Machine Learning techniques.

The system analyzes important factors such as content type, genre, language, release year, ratings, views, popularity, and audience engagement. This helps identify popular content categories, rating patterns, language-wise preferences, and changes in audience interests over time.

The interactive dashboard presents important information through KPI cards, content trends, genre analysis, language distribution, rating charts, popularity analysis, sentiment analysis, and top-content tables. These visualizations make complex OTT datasets easier to understand and support effective content analysis.

The prediction module can estimate content popularity, ratings, or audience engagement using historical data. By analyzing previous content characteristics and performance, the system can provide useful predictive insights that may support content planning and audience analysis.

In the future, the system can be enhanced with real-time viewing analytics, personalized recommendations, content demand forecasting, audience segmentation, churn prediction, advanced sentiment analysis, and real-time trend detection. These improvements can make the framework more intelligent and useful for analyzing OTT content performance and evolving audience preferences.

REFERENCES

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, "Real-Time Object Detection in Drone Surveillance Using YOLOv5," in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, "Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks," in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, "Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology," in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.