International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0046
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 291-298

Transformative Applications of Data Science and Machine Learning: Innovations in Healthcare, Entertainment and Personal Finance

Mallikarjuna Nandi[1], Krishna Manam[2], E susmitha[3], Kavya sri Polampalli[4]
[1,2]Associate professor, Dept. of CSE, Rajiv Gandhi University of Knowledge and Technologies, Ongole, Andhra Pradesh, India.
[3]Associate professor, Dept. of CSE, Rajiv Gandhi University of Knowledge and Technologies, RK Valley Andhra Pradesh, India.
[4]UG Scholar, Dept. of CSE, Rajiv Gandhi University of Knowledge and Technologies, Ongole, Andhra Pradesh, India.
Email ID: *nandimalliap@gmail.com[1], km@rguktong.ac.in[2], susmisuna@gmail.com[3], kavyasai.0921@gmail.com[4]*

Abstract
Data science and machine learning have become transformative tools for addressing challenges across diverse domains. This paper presents three projects that leverage these technologies to deliver innovative solutions in healthcare, entertainment, and personal finance. The first project focuses on heart disease prediction, utilizing machine learning algorithms to analyze key medical parameters, provide early diagnostic insights. By enabling timely interventions, this approach has the potential to significantly improve patient outcomes and alleviate the burden on healthcare systems. The second project delves into viewer analytics for the OTT platform Hotstar. Using advanced data analysis techniques, the project identifies patterns in viewer behavior, including preferences, peak viewing times, and genre popularity. These insights can be instrumental in optimizing content recommendations, enhancing user engagement, and driving business growth in the entertainment industry. The third project introduces a smart expense tracker designed to empower individuals in managing their finances. By employing predictive analytics, the tracker not only categorizes expenses but also forecasts future spending patterns, offering personalized budgeting advice and promoting financial well-being. Collectively, these projects demonstrate the versatility and impact of machine learning and data analytics in addressing real-world problems. By applying cutting-edge methodologies to distinct sectors, the work underscores the far-reaching potential of data-driven innovation in shaping a smarter, more efficient future.

Keywords: *Machine learning, heart disease prediction, healthcare analytics, OTT platforms, Hotstar viewer analysis, data analytics, expense tracking, financial management, predictive analytics, user behavior, personalized recommendations, data-driven innovation.*

1. Introduction

The rapid evolution of data science and machine learning has opened new frontiers in solving complex problems across diverse fields. These technologies provide unprecedented opportunities to analyze large datasets, uncover hidden patterns, and deliver actionable insights, drivinginnovation in healthcare, entertainment, and finance. In thehealthcare sector, the early detection of life-threatening conditions like heart disease remains a significant challenge. Predictive models powered by machine learning can analyze patient data to identify risk factors and provide timely interventions, revolutionizing traditional approaches to diagnosis and treatment. Meanwhile, in the entertainment industry, the rise of OTT platforms has created a demand for personalized user experiences. Platforms like Hotstar generate vast amounts of viewer data, which, when analyzed effectively, can provide insights into audience behavior, optimize content delivery, and enhance user engagement, leading to improved satisfaction and business growth. In personal finance, the increasing complexity of managing expenditures has underscored the need for smarter tools. A predictive expense tracker can provide users with the ability to not only monitor their spending but also forecast future expenses and receive personalized budgeting advice, fostering better financial decisions and habits. This paper presents three independent projects, each addressing challenges in these fields. By leveraging machine learning and data analytics, these solutions demonstrate the adaptability and transformative potential of data-driven innovation across industries. This work presents several machine learning approaches for predicting heart diseases, using data of major health factors from patients. The paper demonstrated four classification methods: Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB), to build the prediction models [1]. Effective decision-making in healthcare and agriculture can be challenging due to the complexity and volume of data involved. Traditional methods often require extensive manual analysis and domain-specific expertise, which can be time-consuming and prone to error. The challenge is compounded by the need for real-time insights and accurate predictions to address critical issues such as disease diagnosis [8][9].

2. Related Work

Heart Disease Prediction: Healthcare industries generate enormous amount of data, so called big data that accommodates hidden knowledge or pattern for decision making. The huge volume of data is used to make decision which is more accurate than intuition. Exploratory Data Analysis (EDA) detects mistakes, finds appropriate data, checks assumptions and determines the correlation among the explanatory variables.[5].Predictive models for heart disease have been widely explored in the healthcare domain, where the focus has been on using machine learning algorithms to analyze patient data and identify risk factors. Xia et al. (2015) explored the use of various classification algorithms, including Decision Trees, Random Forest, and Support Vector Machines (SVM), to predict heart disease risk based on clinical data. These algorithms have shown varying degrees of accuracy and interpretability, making them effective tools for early diagnosis. Kotsiantis et al. (2007) investigated ensemble learning methods, combining multiple models to enhance prediction performance. The Cleveland Heart Disease dataset, introduced by Dua and Graff (2019), has become a benchmark in the field, facilitating the development of predictive models. More recently, Jafari et al. (2020) applied deep learning models, such as neural networks, to detect non-linear patterns in the data, achieving better performance than traditional machine learning methods. These advancements highlight the growing sophistication in the application of machine learning for heart disease prediction. OTT Viewer Analysis: The analysis of viewer behavior on OTT platforms, such as Hotstar, has gained significant attention due to the increasing need for personalized content and engagement strategies. Deo et al. (2020) employed clustering techniques, including K-means and hierarchical clustering, to segment users based on their viewing habits and preferences, thereby improving content recommendations. Their work emphasizes the importance of identifying distinct viewer segments to tailor content offerings effectively. Sengupta and Maity (2019) leveraged sentiment analysis of user reviews, which allowed them to understand how viewers felt about specific genres and content types, providing deeper insights into engagement and satisfaction. In addition, Rathore et al. (2018) explored collaborative filtering and content-based recommendation systems, which have been widely adopted to personalize the user experience on streaming platforms.[6] These studies demonstrate the power of data analytics in optimizing content strategies and increasing user retention on OTT platforms. Expense Tracking and Financial Management: The use of machine learning in financial management, particularly in expense tracking, has been a growing area of research. Chen et al. (2018) applied machine learning algorithms to categorize transaction data and provide users with insights into their spending patterns. This approach allows users to track and manage their finances more efficiently by automating the categorization process. Karlan et

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0046
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 291-298

al. (2016) incorporated principles from behavioral economics to develop financial tools that not only track spending but also offer personalized budgeting advice to users. Their work demonstrated how integrating behavioral insights could lead to better financial decision-making. Bhattacharya et al. (2019) introduced predictive models for forecasting future expenses, allowing users to anticipate their financial needs and adjust their budgets accordingly. Furthermore, Zhou et al. (2020) utilized time-series forecasting and regression models to predict income and expense trends, providing users with a forward-looking view of their financial situation.[4] These approaches highlight the growing potential of machine learning in creating intelligent financial tools that help individuals manage their finances more effectively. Each of these studies contributes to a broader understanding of how data science can be applied to healthcare, entertainment, and finance. This paper builds upon these foundational works by integrating advanced predictive modeling techniques and focusing on specific, actionable outcomes tailored to real-world challenges.

3. Proposed Methodology

The proposed system combines the power of machine learning and data analytics to tackle key challenges across healthcare, entertainment, and personal finance sectors. In the healthcare domain, the system predicts heart disease risk by analyzing patient data, such as age, cholesterol levels, blood pressure, and family history. Machine learning algorithms, including classification models, provide early detection, allowing healthcare professionals to intervene promptly and improve patient outcomes. The system can be integrated into clinical environments, providing real-time risk scores and personalized health recommendations. In the entertainment industry, particularly for OTT platforms like Hotstar, the system gathers and analyzes vast amounts of viewer data, including user preferences, watch history, and demographic information. By utilizing clustering techniques, the system segments viewers based on their viewing habits, and applies recommendation algorithms—such as collaborative filtering and content-based filtering—to provide personalized content suggestions[3]. This enhances user engagement by delivering relevant shows, movies, or genres tailored to individual preferences, thereby improving overall user experience and retention on the platform. For personal finance management, the system tracks users' expenses through integration with their bank accounts or credit cards, categorizing transactions into relevant groups (e.g., groceries, entertainment, bills). Predictive analytics are applied to forecast future spending patterns, helping users anticipate their financial needs and avoid overspending. The system offers budgeting tools that suggest ways to save based on historical spending data, providing real-time alerts when users approach their budget limits. Additionally, it offers financial insights and recommendations, empowering users to make smarter financial decisions, plan better, and ultimately improve their financial health. The architecture of the system includes secure cloud-based data storage, advanced machine learning models, and an intuitive user interface for web or mobile platforms. Each domain's model is continuously updated and refined to ensure accurate predictions and recommendations. The user interface enables seamless interaction, allowing users to input data, receive insights, and take action based on system suggestions. By integrating these technologies, the proposed system delivers a comprehensive, data-driven solution across multiple domains, improving decision-making, optimizing user experiences, and driving positive outcomes in healthcare, entertainment, and finance.

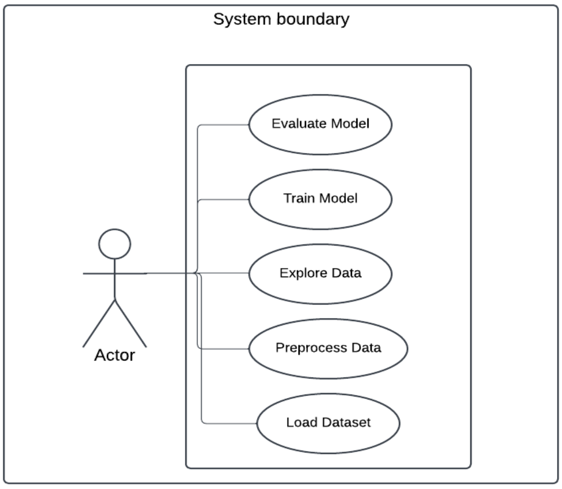

Figure 1 **System Boundry**

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0046
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 291-298

The UML diagram represents a machine learning workflow where the "Actor" interacts with the system to perform key tasks in training and evaluating a model. The process begins with the "Load Dataset" use case, where the actor loads the necessary data into the system, Figure 1.

Table 1 Features in Cleveland Heart Disease Dataset

Table 1.

List of features in the Cleveland heart disease dataset.

Order	Feature	Description	Feature Value Range
1	Age	Age in years	29 to 77
2	Sex	Gender	Value 1 = male Value 0 = female
3	Cp	Chest pain type	Value 0: typical angina Value 1: atypical angina Value 2: non-anginal pain Value 3: asymptomatic
4	Trestbps	Resting blood pressure (in mm Hg on admission to the hospital)	94 to 200
5	Chol	Serum cholesterol in mg/dL	126 to 564
6	Fbs	Fasting blood sugar > 120 mg/dL	Value 1 = true Value 0 = false
7	Restecg	Resting electrocardiographic results	Value 0: Normal Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of >0.05 mV) Value 2: showing probable or definite left ventricular hypertrophy by Estesâ€™ criteria
8	Thalach	Maximum heart rate achieved	71 to 202
9	Exang	Exercise-induced angina	Value 1 = yes Value 0 = no
10	Oldpeak	Stress test depression induced by exercise relative to rest	0 to 6.2
11	Slope	The slope of the peak exercise ST segment	Value 0: upsloping Value 1: flat Value 2: downsloping
12	Ca	Number of major vessels	Number of major vessels (0â€"3) colored by fluoroscopy
13	Thal	Thallium heart rate	Value 0 = normal; Value 1 = fixed defect; Value 2 = reversible defect
14	Target	Diagnosis of heart disease	Value 0 = no disease Value 1 = disease

Next, the actor moves on to the "Preprocess Data" stage, where they clean and transform the data by handling missing values, normalizing features, and encoding categorical variables. Once the data is preprocessed, the actor explores it by visualizing patterns and relationships in the "Explore Data" phase. After this, the actor trains the model using appropriate algorithms in the "Train Model" use case. Finally, the actor evaluates the model's performance in the "Evaluate Model" stage, assessing metrics like accuracy, precision, and recall to determine the model's effectiveness. These steps, encapsulated within the system boundary, represent a typical machine learning pipeline designed for model development and evaluation. Researchers suggest that the five class features of this data set be reduced to two classes; 0 = no disease and 1 = disease. The target feature refers to the presence of heart disease in the subject. Table 1 shows the features included in the Cleveland heart disease dataset. [2]

4. Steps Involved

4.1 Data Collection and Integration

Healthcare: Collect patient data such as medical history, age, cholesterol levels, blood pressure, ECG results, and other relevant health metrics. This can be sourced from hospital records or patient surveys. **Entertainment:** Gather viewer behavior data from Hotstar or other OTT platforms, including watch history, genres, user demographics, device usage, ratings, and feedback.

Finance: Integrate with users' bank accounts or credit card data to track their transactions. This can include expense amounts, categories, and timestamps.

4.2 Data Preprocessing and Cleaning

Healthcare: Clean the medical data by handling missing values, normalizing numerical features, and encoding categorical data (e.g., gender, smoking habits). Ensure the data is suitable for machine learning algorithms.

Entertainment: Preprocess viewer data by removing duplicates, handling missing data, and standardizing time stamps and viewer ratings. **Finance:** Clean and categorize transaction data by converting raw transaction details into meaningful categories such as food, transportation, utilities, etc. Remove any erroneous or incomplete records.

4.3 Feature Selection and Engineering

Healthcare: Identify the most important features for predicting heart disease using techniques like Recursive Feature Elimination (RFE) or Principal Component Analysis (PCA).

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0046
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 291-298

Entertainment: Perform feature engineering to identify relevant viewer behaviors and preferences, such as frequency of viewing, preferred genres, and time spent watching specific content.

Finance: Create meaningful features from financial data, such as average monthly spending, highest expenditure categories, or frequency of expense spikes.

4.4 Model Evaluation and Tuning

Healthcare: Evaluate the models using metrics such as accuracy, precision, recall, and F1-score. Tune hyperparameters using GridSearch or RandomSearch to optimize performance.

Entertainment: Evaluate clustering results using silhouette scores and check recommendation accuracy with metrics like Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE). Fine-tune the recommendation algorithm.

Finance: Assess prediction accuracy for expense forecasting using metrics like Mean Absolute Error (MAE) and R2. Refine the model by adjusting parameters and validating predictions with test data.

4.5 Real-Time Data Processing and Feedback Loop

Healthcare: Continuously update the model with new patient data to refine predictions and improve accuracy. This may involve retraining the model periodically.

Entertainment: Continuously gather and analyze viewer data to adapt and refine content recommendations based on changing viewing habits or preferences.

Finance: Continuously monitor users' spending behaviors and update predictive models to reflect new financial patterns. Alerts and suggestions can be adjusted based on new data.

4.6 Monitoring and Maintenance

Healthcare: Continuously monitor the accuracy of the heart disease prediction model and update it as more data is collected. Implement system updates to ensure it remains aligned with the latest research.

Entertainment: Track the performance of the recommendation engine and user satisfaction to ensure content suggestions are improving engagement. Update algorithms as viewer preferences evolve.

Finance: Monitor the performance of the expense tracker by gathering user feedback and analyzing predictive accuracy. Periodically update the forecasting models to account for changes in financial trends.

4.7 Scalability and Future Enhancements

Healthcare: Explore integration with other health data sources (e.g., wearable devices, hospital systems) to enhance the prediction model and make the system more comprehensive.

Entertainment: Expand the system to support multiple OTT platforms and integrate advanced NLP techniques for sentiment analysis to refine content recommendations.

Finance: Add features for investment forecasting, savings advice, or debt management. Enhance the model by incorporating more granular financial data for more personalized recommendations.

5. Results

The results of the machine learning workflow, as depicted in the UML diagram, would typically be as follows:

- **Loaded Dataset:** The actor successfully loads the dataset, ensuring that the data required for the model is available and in an accessible format.
- **Preprocessed Data:** After preprocessing, the data is clean, transformed, and ready for analysis. This means that missing values are imputed or removed, categorical variables are encoded, and numerical data is normalized or scaled if necessary. The dataset is now suitable for feeding into machine learning models.
- **Explored Data:** In this phase, the actor would gain insights into the data through visualizations or statistical analysis. The results might reveal trends, outliers, correlations, and distributions within the data, helping the actor make informed decisions on which features to focus on during model training.
- **Trained Model:** The machine learning model is trained on the preprocessed and explored data. The result is a model that has learned patterns from the data, with parameters adjusted to optimize performance. This could be a classification, regression, or clustering model, depending on the problem being solved.
- **Evaluated Model:** The model's performance is evaluated using appropriate metrics (e.g., accuracy, precision, recall, F1-score, or mean squared

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0046
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 291-298

error for regression tasks). The results would indicate how well the model generalizes to unseen data. If the model's performance is satisfactory, it can be deployed; otherwise, the actor may iterate through the process by adjusting features, tuning hyperparameters, or choosing a different model. [7]

These results Figure 2 to Figure 13 demonstrate the effectiveness of the machine learning pipeline and provide insights into areas for improvement or further refinement in the model's development.

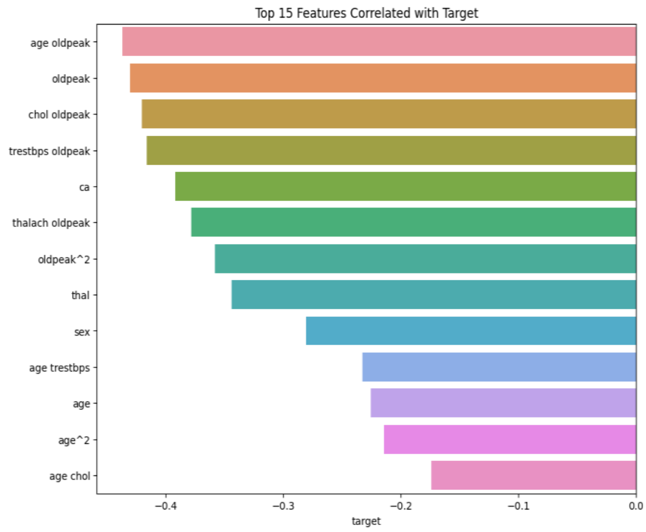

Figure 2 Compute the Correlation Matrix, Plotting the Top 15 Features Most Correlated with The Target Variable

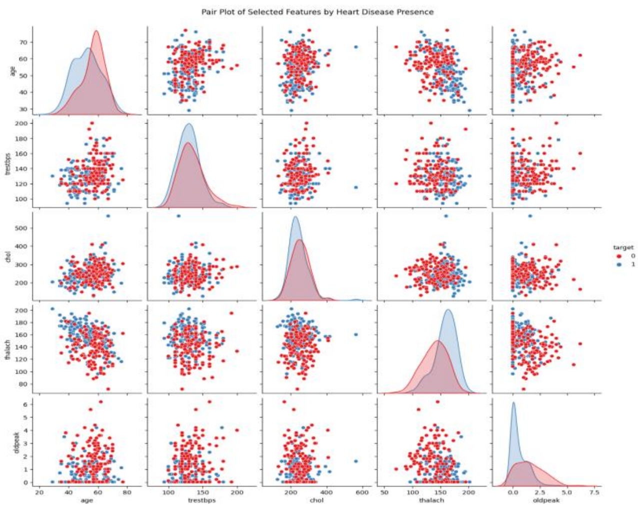

Figure 3 Pair Plot of Selected Features by Target

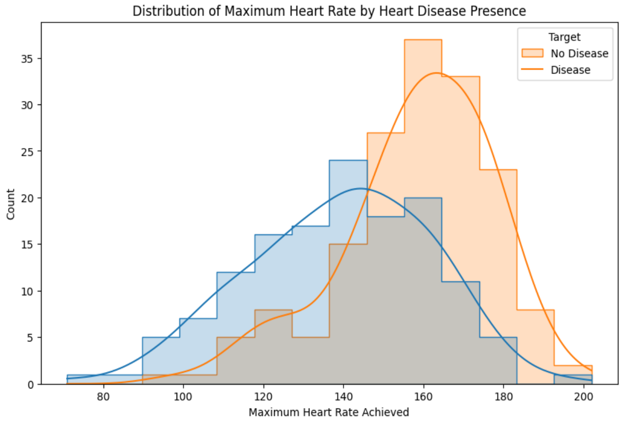

Figure 4 Distribution Plot for Max. Heart Rate Achieved by Target

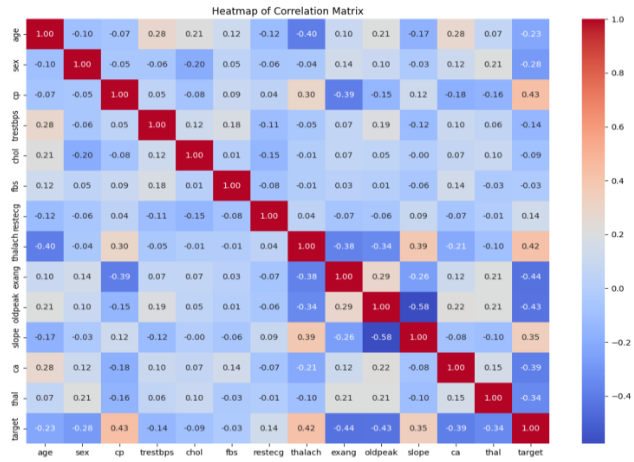

Figure 5 Computing the Correlation Matrix, Plotting Heat Map of The Correlation Matrix

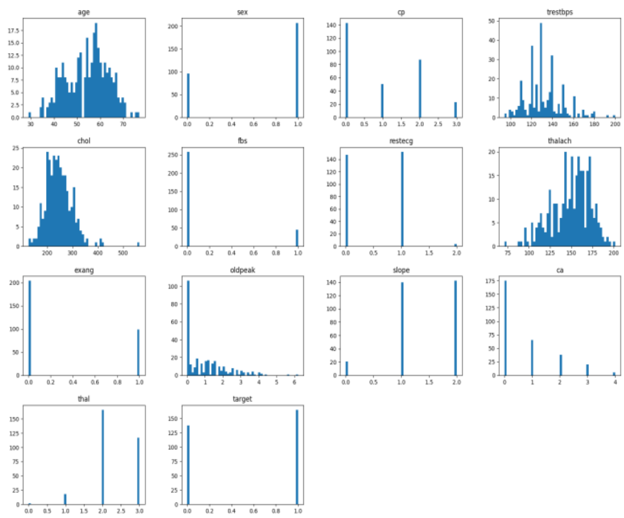

Figure 6 Histogram of Each Attribute

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0046
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 291-298

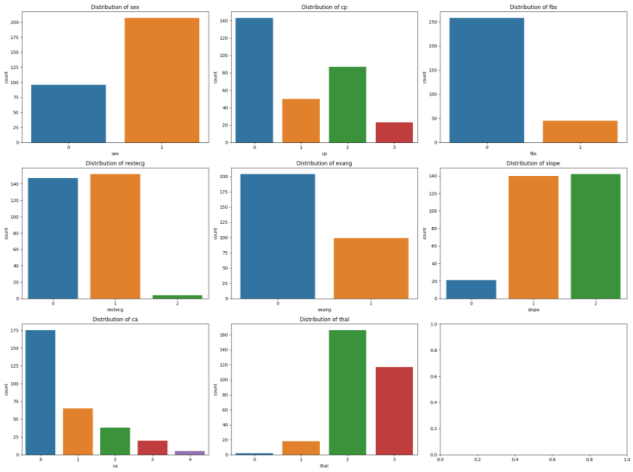

Figure 7 Plotting Categorical Features

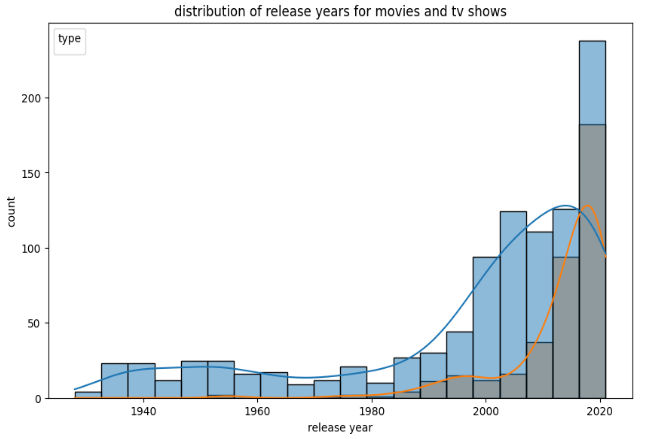

Figure 10 Distribution of Release Years for Movies and Tv Shows

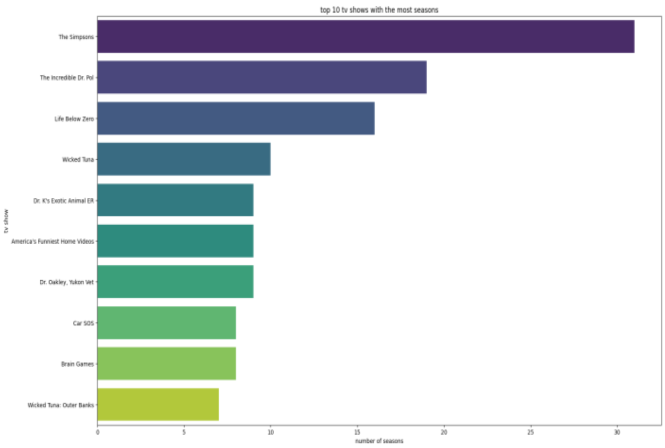

Figure 8 Getting Top Actor/Actresses by Frequency

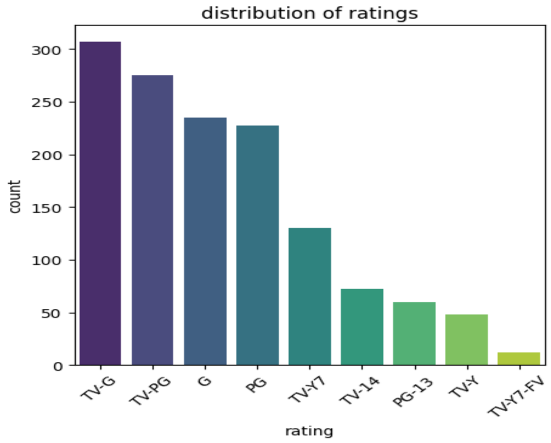

Figure 11 Distribution of Ratings of The Movies

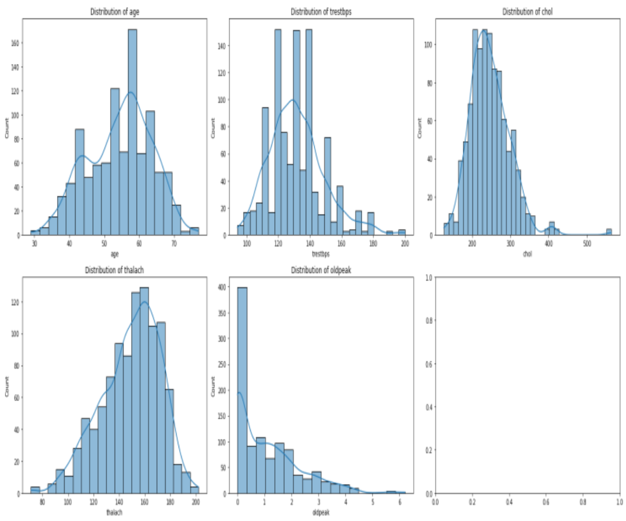

Figure 9 Plotting Continuous Features

```
 Enter Transaction Details:
Description: Expense tracker
Category: Income
Amount: 15000
Type (Income or Expense): Income
 Would you like to add another transaction?   (YES / NO):YES

 Enter Transaction Details:
Description: Expenses
Category: Shopping
Amount: 2000
Type (Income or Expense): Expense
 Would you like to add another transaction?   (YES / NO):YES

 Enter Transaction Details:
Description: Expense
Category: Rent
Amount: 6000
Type (Income or Expense): Expense
 Would you like to add another transaction?   (YES / NO):YES

 Enter Transaction Details:
Description: Expense
Category: Food
Amount: 3000
Type (Income or Expense): Expense
 Would you like to add another transaction?   (YES / NO):YES

 Enter Transaction Details:
Description: Transportation
```

Figure 12 Displays Income, Expense and Balance and Transaction Details

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0046
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 291-298

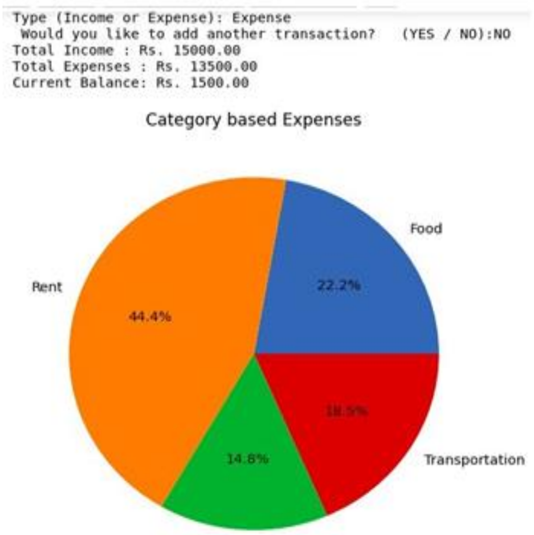

Figure 13 Visualization of Expenses is Shown by Means of a Pie Chart by Diving Categories

References

[1]. Y. Anitha, R. Ranjini, S. Gomathi, “Easy App for Expanses Manager Using Android,” International Journals of Computer Techniques, Volume 3, Issue 2, ISSN: 2394-2231, March-April 2016.
[2]. Adepegba, Oluwafunmilola, Fayemiwo, Michael, Oduwole, Oludayo, & Onamade, A.A., “An Android-Based Mobile Application for Tracking Daily Expenses,” 10.2/AIMS/iSTEAMS-019/V21N1P9, 2019.
[3]. Thanapal, M. P., Patel, Y., Lokesh, R. T. P., & Satheesh, K. J., “Income and Expense Tracker,” Indian Journal of Science and Technology, 8(S2), pp. 118-122, 2015.
[4]. Chaimaa Boukhatem, “Department of Electrical Engineering, University of Sharjah,” Conference on 21-24 February 2022, Added to IEEE Xplore on 18 March 2022.
[5]. Predictive data mining for medical diagnosis: “An Overview of Heart Disease Prediction,” International Journal of Computer Applications, Volume 17, Issue 8, pp. 43-48, 2011.
[6]. Chen, A. H., Huang, S. Y., Hong, P. S., Cheng, & Lin, E. J., “HDPS: Heart Disease Prediction System,” 2011 Computing in Cardiology, pp. 557-560, IEEE, September 2011.
[7]. Raihan, M., Mondal, S., More, A., Sagor, M. O. F., Sikder, G., Majumder, M. A., & Ghosh, K., “Smartphone-Based Ischemic Heart Disease (Heart Attack) Risk Prediction Using Clinical Data and Data Mining Approaches: A Prototype Design,” 2016 19th International Conference on Computer and Information Technology (ICCIT), pp. 299-303, December 2016.
[8]. Nandi, M., Vinitha, M., Nagaraja Naik, B., Yedukondalu Naik, B., &Dhramateja, G. (2024). Omnicare: A comprehensive ML/DL-based prediction system for healthcare and agriculture. IRJASH, 6(10), 313–319. https://doi.org/10.47392/IRJASH.2024.041
[9]. Vinitha, M., & Nandi, M. (2024). Plant leaf disease detection using convolutional neural network. International Journal of Science and Research (IJSR), 13(4), 1545–1548. https://doi.org/10.21275/SR24422151104