Journal of Computer Science

Research Article

Exploratory Data Analysis of Applications of Encryption for Netflix by ML Models

Ananth Maria Pushpa and Subramanian Shanmugam

Department of Mathematics, PRIST Deemed to be University, Thanjavur, India

Article history
Received: 19-12-2024
Revised: 26-02-2025
Accepted: 21-03-2025

Corresponding Author:
Ananth Maria Pushpa
Department of Mathematics, PRIST Deemed to be University, Thanjavur, India
Email: mariapushpa2017@gmail.com

Abstract: This study examines the performance of various machine learning algorithms in analyzing Netflix's dataset, with a focus on encryption methods used for data protection. We evaluated several models including Ada BoostM1, IBk, Random Forest, and Decision Stump using multiple performance metrics. This work analysis revealed that the Ada BoostM1 model achieved the highest accuracy at 97.14%, outperforming other algorithms. It also demonstrated superior performance in precision (0.97), recall (0.97), and F-Score (0.97). The Random Forest algorithm showed the highest ROC value of 0.98. In contrast, the Decision Stump algorithm consistently underperformed, showing the lowest precision (0.71), recall (0.71), F-Score (0.70), and ROC value (0.64). The IBk model also showed relatively low accuracy at 88.09%. Here evaluated the models using the kappa coefficient and Matthews Correlation Coefficient (MCC). Ada BoostM1 achieved the highest scores in both metrics (0.94 for kappa and MCC), while Decision Stump showed the lowest (0.40 for kappa and 0.41 for MCC). Our findings suggest that Ada BoostM1 and Random Forest algorithms are the most effective for analyzing Netflix's dataset, potentially offering insights into the company's competitive strategy and development model. This research contributes to understanding the application of machine learning in analyzing streaming service data and the effectiveness of various algorithms in this context.

Keywords: Decision Stump, Ada Boost, ROC, PRC, Advanced Encryption Standard, Netflix, Random Forest

Introduction

The use of technology has become ubiquitous in modern society. Due to AI, jobs that were once subject to strict regulations are now thriving. first to third Unquestionably, the Internet of Things has given people the ability to do incredible things. E-commerce behemoth Amazon created Alexa, a cloud-based speech service that provides users with an easy way to interact with technology through Automatic Speech Recognition. By making it accessible to people of all ages, the extraordinary technique of transcribing spoken words into text has considerably increased the technology's usefulness. Artificial Intelligence (AI) describes how computers and other technological systems can mimic human intellect. Natural Language Processing, computer vision, voice recognition and expert systems are some of the most common areas where AI is used extensively. AI is more than just a computer program that can mimic human brainpower, according to research by Stephen Dick. On the contrary, it emphasizes how our definition of intelligence has evolved. Taking action is the defining characteristic that distinguishes AI from general-purpose software. AI enables computers to react autonomously to environmental cues, even when programmers do not directly influence or anticipate these cues. In references. During the COVID-19 era, innovations occurred that were unthinkable in any other era. One example is the format of online education lessons and exams, as well as remote work from home. In 1927, having access to electronic entertainment became even more crucial during the closure. From 2019 through 2021, there was massive expansion for the current OTT IPTV (Internet Protocol Television) platforms, so they might thrive in the entertainment sector. People of all ages were able to pass the time while movie theaters were closed by watching content on various over-the-top (OTT) platforms. In, with the help of OTT services, people were able to watch movies, TV shows, and original material online.

With OTT media services, content is distributed and sent using the internet instead of traditional means like cable or satellite. This is what led to the proliferation of streaming media. You may find Verizon's definition of streaming on their website. The term streaming is defined as any media content, live or recorded, supplied

via the internet and played back in real-time to computers and mobile devices. One kind of content that may be viewed through streaming services is music videos, webcasts, movies, and television series. The advent of more widely available and abundantly available OTT IPTV services has revolutionized the media environment in every other way imaginable. The skyrocketing popularity of OTT IPTV services has ignited a content arms race among content providers. Quality of service experience (QoSE) and customer engagement (CE) are two other factors that have been discussed in publications by different scholars. There are many more factors to consider, such as the customer's attitude, the product's usefulness and quality, the ease of use, the perceived risk, the security measures, the degree of involvement, and the entire service experience.

This study aims to conduct a comprehensive analysis of Netflix's ecosystem through three key objectives. First, we will perform an in-depth Exploratory Data Analysis (EDA) of Netflix data to uncover patterns in content attributes, user interactions, and viewing behaviors. Second, we will investigate the role of advanced encryption algorithms in safeguarding this sensitive user data. Finally, the study will evaluate the impact of various machine learning models on the accuracy and personalization of the platform's recommendation system.

Literature Review

A key factor in Netflix's market dominance is its strategic use of technology to enhance user experience and retention. The platform's market leadership is evidenced by its popularity; a survey of 307 OTT users in Delhi found that 63.8% preferred Netflix as their primary streaming service (Rastogi *et al.*, 2023). This preference is not accidental but is driven by a sophisticated technological backbone. A central component is the Algorithm for Video Attribution (AVA), a tool designed to optimize performance by efficiently selecting and delivering visual content, thereby improving system efficiency and user engagement (Rastogi *et al.*, 2023).

Netflix's success is also a case study in corporate reinvention and scaling entrepreneurial spirit. The company exemplifies Schumpeter's concept of creative destruction, having repeatedly disrupted the media landscape by building new business models on the remnants of old ones, such as transitioning from DVD rentals to streaming (Mier & Kohli, 2021). A critical research question involves understanding how Netflix has managed to scale to over 9,000 employees while maintaining its innovative and agile startup mentality, a challenge many growing companies fail to overcome (Mier & Kohli, 2021).

Central to this technological edge are the advanced recommender systems that personalize the user experience. Netflix was a pioneer in leveraging artificial intelligence (AI) to power its recommendation engine, which analyzes user data, including watch history, searches, and ratings, to suggest relevant content and maximize long-term engagement (Gomez-Uribe & Hunt, 2016). While Netflix employs a complex ecosystem of machine learning models, its competitors utilize similarly advanced techniques. For instance, Amazon Prime Video uses Graph Convolutional Networks (GCNs) to integrate data from various sources about videos and users (Arti, 2021). Additionally, Amazon's X-Ray feature uses the Amazon Recognition API to perform image identification on video frames, linking recognized faces to IMDb data to provide viewers with enriched contextual information. This demonstrates how both leading platforms heavily rely on data science and deep learning to facilitate content discovery and maintain user interest.

A significant challenge in developing effective recommender systems is data sparsity, which occurs when there is insufficient information on user-item interactions to generate accurate suggestions. A common technique to address this in collaborative filtering (CF) is Matrix Factorization (MF), which aims to decompose the user-item interaction matrix into lower-dimensional latent factors to learn hidden representations for users and items (Behera & Nain, 2022). Building upon this, Behera and Nain (2022) propose a Deep Non-Negative Matrix Factorization (DNNMF) model. This advanced method enhances traditional MF by integrating deep neural networks and applying non-negative constraints to the embedding layers. This combination allows the model to capture complex, non-linear user-item relationships while ensuring that the learned feature vectors for users and items are non-negative, leading to more interpretable factors. Empirical tests demonstrate that the DNNMF model robustly addresses the sparsity problem in CF, outperforming newer models and maintaining efficacy across varying levels of data sparsity, thereby presenting a viable solution for improving recommendation accuracy in real-world applications.

The rise of internet-based television services like Netflix has fundamentally reshaped viewing habits, most notably through the phenomenon of binge-watching, the consecutive viewing of multiple episodes of a series. Castro et al. (2021) conducted a study to explore the motivations and emotional impacts of this behavior, analyzing 40 Netflix sessions from 11 millennials using a novel mixed-methods approach that combined objective data from a browser extension with subjective pre- and post-viewing questionnaires. Their findings indicate that binge-watching is primarily a solitary, evening activity used for unwinding, escapism, and coping, with an average session lasting about two hours and ten minutes. The research also revealed that this activity modulates viewers' affective states; for instance, watching a comedy reduced negative affect, while watching a drama slightly increased it, though overall arousal levels remained stable. This study significantly contributes to

understanding the psychological drivers and emotional consequences of engaging with serialized television fiction in the streaming era.

While algorithmic recommender systems are central to media platforms like Netflix, their role and impact are subjects of critical scholarly debate. Frey (2021) provides a counter-narrative to the popular perception of these systems as entirely novel and universally beneficial. His work questions the revolutionary status often ascribed to recommender systems, arguing they are neither as transformative nor as threatening as commonly portrayed. Frey (2021) posits that despite the power of AI-driven algorithms, human curation, expert opinion, and social recommendations continue to play a vital role in how viewers discover content. The book challenges the "big data myth" perpetuated by Netflix, demonstrating that audience choices are influenced by a complex mix of algorithmic and non-algorithmic sources, and calls for more empirical research into the actual decision-making processes of contemporary viewers.

From an engineering perspective, Netflix's recommender system is a complex and evolving business-critical tool. As Gomez-Uribe and Hunt (2021) detail, it is not a single algorithm but an extensive ecosystem that includes both recommendation and search algorithms. To maintain a competitive edge with a massive global user base, Netflix relies heavily on A/B testing to iteratively refine its system, with a primary focus on long-term member retention and engagement (Gomez-Uribe & Hunt, 2021). While online A/B testing is invaluable, the authors note that offline testing using historical interaction data is also a crucial component of their development cycle. However, this approach presents challenges, particularly in the design and interpretation of A/B tests. In response to these complexities and the platform's global nature, the recommender engine is continuously updated, with recent developments focusing on becoming more globally aware and sensitive to linguistic nuances.

The underlying architecture of this system is a hybrid model that integrates multiple approaches for optimal performance. Collaborative filtering forms one pillar, analyzing user behavior and preferences to surface content enjoyed by other users with similar tastes (Gomez-Uribe & Hunt, 2021). The second pillar is content-based filtering, which recommends new titles based on the attributes of content a user has previously enjoyed. By combining these two approaches into a hybrid system, Netflix creates a more robust and accurate recommendation engine. The platform is increasingly leveraging deep learning techniques, utilizing specialized models for different types of suggestions, which has significantly enhanced the precision of its personalization.

Ultimately, every choice a user makes contributes to a unique profile, allowing Netflix to tailor the experience based on individual habits, viewing patterns, and interactions with the service. This intricate synthesis of collaborative and content-based filtering, powered by advanced machine learning and continuous interface optimization, ensures that each user's homepage is a highly individualized portal designed to maximize engagement and satisfaction.

Research into the societal impact of Netflix extends beyond its algorithms to the content it disseminates, particularly regarding the representation of social groups and health behaviors. A comparative content analysis by Çelik (2023) examined the portrayal of older adults in eight popular Turkish prime-time TV shows and eight Netflix Turkish originals. The study found a significant underrepresentation of older characters on both platforms relative to their actual demographic in Turkey, suggesting a pattern of exclusion and age-based discrimination. While Netflix originals featured more realistic portrayals of older people than their prime-time counterparts, this difference was not statistically significant, and both platforms shared several critical shortcomings. Notably, older women were severely underrepresented compared to older men, and no older character held a main protagonist role in any series on either service. Furthermore, both platforms perpetuated negative stereotypes, particularly concerning the employability of older individuals, a bias that placed older women in a state of "double jeopardy" due to the intersection of age and gender.

Separate from representation, the prevalence of unhealthy imagery in Netflix's content library raises public health concerns. Hanewinkel et al. (2023) analyzed the frequency of smoking in Netflix films and the effectiveness of age-based ratings in Germany and the United States. Their findings indicate that smoking is a common occurrence in Netflix films. More critically, the study concluded that Netflix does not adequately adhere to the World Health Organization's Framework Convention on Tobacco Control, which recommends restricting youth exposure to smoking imagery. The research highlighted a disparity in youth protection between the two countries; while the U.S. rating system was more effective, it still deemed some films containing smoking scenes appropriate for youth audiences. In contrast, Germany's less stringent system rated half of the movies containing smoking scenes as acceptable for children, indicating a significant failure in regulatory alignment with public health guidelines.

The development of sophisticated recommender systems that leverage social connections to enhance user experience presents a significant area of innovation. Khalique et al. (2022) address this challenge by proposing a deterministic model for Over-the-Top (OTT) platforms that quantifies the degree of friendship between users based on their mutual content likings and recommendations. This hybrid recommendation system analyzes shared preferences to infer social ties, creating a

more nuanced understanding of user relationships. The proposed model demonstrates efficacy in mapping viewer affinities, thereby offering a pathway to a more socially-aware and engaging user experience.

In a different domain, the integration of machine learning into healthcare, particularly electronic triage (E-triage) and remote patient priority systems, represents a critical advancement. Salman et al. (2021) conducted a comprehensive review that synthesizes research on sensor-based telemedicine systems utilizing machine learning algorithms. Their work establishes a novel crossover taxonomy to systematically map specific machine learning algorithms to their corresponding applications in telemedicine, which includes both synchronous (real-time) and asynchronous (store-and-forward) modalities. This taxonomy provides a structured overview of the relationships between machine learning techniques, input data, performance metrics, and clinical outcomes.

The impetus for this technological integration is the growing global burden of chronic diseases, such as diabetes and hypertension, which necessitates intelligent health monitoring systems (Salman et al., 2021). Machine learning offers the potential to revolutionize E-triage and patient prioritization by enabling more accurate and efficient assessment. The review by Salman et al. (2021) not only catalogs the current state of the art but also identifies open research challenges, delineates the benefits of machine learning in telehealth, and proposes directions for future work. This synthesis is designed to foster interdisciplinary collaboration, guiding experts on how to refine AI and machine learning applications to modernize and enhance healthcare delivery systems for the future.

Understanding the prevalence of e-cigarette imagery in popular media is crucial, particularly in content that shapes the perceptions of young adults. Allem et al. (2022) conducted a content analysis to quantify the instances of e-cigarette-related visuals and dialogue in Netflix scripted television and movies popular among viewers aged 18-24. The study identified the most-watched Netflix original content in the U.S. within this demographic from June 2020 to May 2021 using Nielsen ratings, resulting in a sample of 12 movies and 113 episodes from 12 television series.

The methodology involved a rigorous coding process where three coders analyzed over 101 hours of content for e-cigarette appearances, usage frequency, character demographics, brand visibility, and any dialogue about vaping. To ensure reliability, 20% of the content was double-coded. The findings revealed that 13% of the titles (15 out of 125) contained e-cigarette-related content. This included visual depictions of characters holding e-cigarettes in ten titles and vaping-related dialogue without visual props in three titles. In total, e-cigarettes appeared on screen for more than 300 seconds, with an average of 31 seconds of visibility per occurrence, and a character was shown holding an e-cigarette in nearly every instance.

Based on these findings, Allem et al. (2022) conclude that the normalization of e-cigarette use in widely consumed Netflix programming is a significant public health concern. The study underscores an urgent need for targeted health education initiatives to counter the normalization of vaping, particularly among young adults and other vulnerable populations who are heavily influenced by media representations.

The pervasive presence of imagery depicting alcohol, tobacco, and foods high in fat, sugar, and salt (HFSS) in media is a significant factor influencing the consumption behaviors of young people. Research indicates that video-on-demand (VOD) services may contain even more of this content than traditional television programming. A study by Alfayad et al. (2022) sought to quantify this exposure by analyzing images of alcohol, tobacco, and HFSS items in original films from Netflix and Amazon Prime Instant Video accessible in the UK.

The researchers conducted a content analysis of eleven original films from each service released in 2017. The coding process involved assessing five-minute intervals for the presence of the targeted imagery. Out of the 479 intervals coded, the results revealed a high frequency of exposure: alcohol content appeared in 41.7% of intervals (n=200), tobacco in 26.9% (n=129), and HFSS content in 35.24% (n=169). Statistical analysis using chi-square tests showed that the prevalence rates were consistent across both Netflix and Amazon Prime Instant Video films. Furthermore, the occurrence of this content was not correlated with the age ratings of the films, meaning it appeared just as often in content accessible to younger audiences.

Alfayad et al. (2022) conclude that the original films featured on VOD services in the UK frequently contain imagery of alcohol, tobacco, and HFSS products that is likely to promote their use among young viewers. The study highlights a regulatory gap and calls for further research to inform policies that would restrict the display of such harmful and unwanted content on VOD platforms to better protect viewers.

Exposure to alcohol and tobacco imagery in media is a known risk factor that increases the likelihood of use among adolescents and young adults. While previous research has quantified this content in traditional UK television and film, a significant gap existed regarding its prevalence in video-on-demand (VOD) services like Netflix and Amazon Prime. Furthermore, it was unclear whether content differed between services operating under different national jurisdictions, such as Netflix (governed by Dutch law) and Amazon Prime (governed by UK law). To address this, Barker et al. (2019) conducted a content analysis to quantify alcohol and tobacco imagery in the 50 most popular episodes of original programming from 2016 on these two platforms.

The study methodology involved coding 2,704 five-minute intervals for the presence of branding, actual use, implied use, and related paraphernalia for both substances. The findings revealed a substantial presence of this content: tobacco imagery appeared in 13% of the coded intervals (353 out of 2,704) across 74% of the episodes (37 out of 50). Alcohol imagery was even more prevalent, appearing in 13% of intervals (363 out of 2,704) across 94% of the episodes (47 out of 50). While there was no significant difference in prevalence between the two VOD services, a comparison with earlier research on UK prime-time television showed that VOD original programming contained significantly more episodes featuring alcohol and tobacco.

This research demonstrates that VOD original series frequently contain audiovisual depictions of tobacco and alcohol, providing a potent new channel for youth exposure to pro-smoking and pro-drinking stimuli (Barker et al., 2019). This holds true regardless of the service's national base, indicating a globalized media issue. Given the worldwide reach of VOD platforms, this content represents a potentially major, and largely unregulated, vector for the normalization of substance use. This poses a direct challenge to public health guidelines, such as those from the World Health Organization, which recommend that youth should not be exposed to smoking imagery. The shift towards online viewing, accelerated by the COVID-19 pandemic, underscores the urgent need to address these new challenges in youth protection.

Hanewinkel et al. (2023) conducted a content analysis of 235 Netflix original movies from 2021 and 2022 to assess the prevalence of on-screen smoking and the effectiveness of age-based ratings in protecting youth from such imagery. The study aimed to determine: (1) the proportion of smoke-free movies in the sample, (2) the frequency of smoking scenes, and (3) how many movies containing smoking were rated as appropriate for youth in Germany and the United States. For the purpose of this analysis, any film with an age rating of 16 or younger was classified as suitable for children and adolescents.

The findings revealed that smoking is a common occurrence in Netflix's original film library. Out of the 235 movies analyzed, 113 (48.1%) contained smoking scenes. A stark disparity was observed in how these films were rated for youth audiences in the two countries. Of the 113 movies with smoking scenes, only 26 (23.1%) were classified as youth films in the U.S., compared to 57 (50.4%) in Germany, a difference that was statistically significant ($p<0.001$). In total, 3,310 individual smoking incidents were documented. A substantial portion of these, 39.4% (n=1,303), appeared in movies rated for youth in Germany, while 15.8% (n=524) were in U.S. youth-rated films.

Hanewinkel et al. (2023) conclude that Netflix is not adhering to the World Health Organization's Framework Convention on Tobacco Control, which recommends restricting youth exposure to movies containing smoking. While the U.S. rating system offers better protection for children by deeming fewer smoking-containing films appropriate for them, it is not fully effective, as some films with smoking scenes are still accessible to youth. In contrast, Germany's rating system was found to be significantly less protective, with half of all movies containing smoking scenes being deemed acceptable for children.

Materials And Methods

The analysis is conducted using the "Netflix All Weeks Data" collection, a publicly available dataset sourced from the Kaggle repository. This dataset comprises 233,200 instances characterized by 8 attributes, providing a substantial foundation for model training; the specific attributes are detailed in Table 1. The considerable size of this dataset is crucial for developing reliable machine learning models, as it helps mitigate the risk of inaccurate performance metrics that often plague analyses with limited data.

Table 1: Performance of Selected Models

No.	Attribute Name	Data Type
1	country_name	Char
2	country_iso2	Char
3	week	Date
4	category	Char
5	weekly_rank	Numeric
6	show_title	Char
7	season_title	Char
8	cumulative_weeks_in_top_10	Numeric

However, while a large dataset is necessary, it is not a sufficient guarantee of a model's generalizability. A primary concern in machine learning is the bias-variance tradeoff, where a model may either overfit by memorizing noise in the training data (high variance) or underfit by failing to capture underlying patterns (high bias). To address this and bolster the validity of our findings, the study employs k-fold cross-validation. This technique, which iteratively assesses model performance on multiple data subsets rather than a single 90-10 train-test split, provides a more robust and reliable estimate of how the model will perform on unseen data.

Acknowledging the study's scope, we note that a formal bias-variance decomposition or learning curve analysis, which would involve evaluating models on progressively larger data subsets to diagnose overfitting/underfitting and determine the value of additional data, was not conducted. Consequently, while the reported accuracy is a strong indicator of performance, its absolute generalizability to other datasets cannot be conclusively confirmed without further external validation.

To mitigate overfitting, several regularization techniques were employed, tailored to specific model

architectures. These included pruning for decision trees, limiting maximum depth in Random Forests, and implementing dropout within deep learning structures. These methods explicitly penalize model complexity to prevent the learning of spurious patterns in the training data.

Furthermore, hyperparameters were systematically optimized using a combination of Grid Search and Bayesian Optimization. This process aims to minimize both systematic error (bias) and random error (variance), leading to more accurate and stable predictions.

The most definitive assessment of model robustness comes from external validation, where the final model is evaluated on a completely independent, unseen dataset. While this study does not include such external validation, model robustness is justified through a rigorous internal process. This combines the outcomes of k-fold cross-validation with optimal hyperparameter selection and the deliberate restraint of model complexity via the aforementioned regularization techniques.

It is acknowledged that an explicit learning curve analysis and validation on a separate dataset would further strengthen the reliability of the conclusions. These remain valuable directions for future work to conclusively demonstrate the model's generalizability.

EDANetflix Algorithm

The proposed methodology, designated as the EDANetflix framework, is a systematic, multi-stage process for predicting user preferences and content popularity on the Netflix platform. The workflow is outlined below and illustrated in Figure 1.

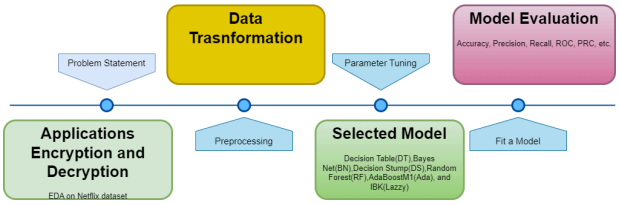

Fig. 1: Proposed system

Step 1: Problem Identification

Define the core objective: to build a predictive model for user preferences or content popularity within the context of the Netflix ecosystem.

Step 2: Exploratory Data Analysis (EDA)

1. Conduct EDA on the Netflix Dataset (D)
2. Uncover underlying trends, patterns, and correlations within the data (T)
3. Identify missing values and outlier (I).

Step 3: Data Collection

1. Dataset is sourced from the Netflix All Weeks Data collection on Kaggle
2. The dataset comprises 233,200 instances (D = $\{x_1, x_2, \ldots, x_{233200}\}$), each described by 8 attributes ($Y_1, Y_2, \ldots, Y_8$) related to movies and TV shows.

Step 4: Data Preprocessing

This stage involves transforming the raw data into a clean, analysis-ready format through a series of operations:

1. Handle Missing Values: Impute or remove missing data points, resulting in an interim dataset (D').
2. Address Outliers: Remove or adjust extreme values to mitigate their influence, yielding dataset (D'').
3. Normalize/Scale Features: Standardize numerical features to a common scale to ensure stable model training, producing dataset (D''').
4. Encode Categorical Variables: Convert categorical attributes into numerical representations, finalizing the preprocessed dataset ('''').

Step 5: Model Selection

1. Candidate Models: A diverse set of machine learning models is selected for evaluation: M = {Decision Tree (DT), Bayesian Network (BN), Decision Stump (DS), Random Forest (RF), AdaBoost, IBk (K-Nearest Neighbors)}.
2. Parameter Tuning: Hyperparameters for each candidate model are optimized to maximize performance.

Step 6: Model Comparison & Evaluation

1. Evaluation Metrics: Model performance is rigorously assessed using a standard set of metrics: E = {Accuracy, Precision, Recall, F1-Score}.
2. Validation: k-Fold cross-validation is employed to ensure the models generalize well to unseen data and to prevent overfitting.

Step 7: Model Fitting

1. The best-performing model from the previous stage is selected (M_selected).
2. This model is then trained on the entire preprocessed dataset (D''''), with final hyperparameter fine-tuning if necessary.

Step 8: Output: Efficient Model (EM)

The final output is a trained and validated Efficient Model: EM = M_selected(D''''). This model is capable of predicting user preferences or content popularity based on the attributes from the Netflix dataset.

Machine Learning Algorithms

To systematically identify the optimal predictive model, a diverse suite of machine learning algorithms was implemented and evaluated on the dataset. This comprehensive approach facilitates robust comparison of

model performance and generalizability across different algorithmic paradigms.

Decision Table

A Decision Table is a rule-based model that functions as a lookup table, making predictions by matching new instances to the most frequent class observed in the training data for specific feature combinations. The algorithm optimizes performance through feature selection to identify the most informative attribute subset, thereby balancing predictive accuracy against model complexity.

Formally, given a dataset $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ where, $x_i \in X$ represents feature vectors and $y_i \in Y$ denotes class labels, a feature subset $S \subseteq X$ is selected.

The decision table mapping is defined as:

$T : S \to Y$

For a new instance x^*, the prediction is:

$\hat{y} = T(x^*[S]) \quad \text{if } \exists s \in S : x^*[S] = s$

Otherwise, a default majority class is assigned:

$\hat{y} = \arg\max_{y \in Y} |\{i : y_i = y\}|$

Feature selection is performed through:

$S^* = \arg\max_{S \subseteq X} \text{Score}(S, D)$

where Score represents an evaluation metric. The algorithm jointly optimizes $|S|$ and $|T|$ to achieve an optimal balance between model complexity and predictive accuracy.

Bayesian Network

A Bayesian Network (Bayes Net) is a probabilistic graphical model that represents a set of random variables and their conditional dependencies through a directed acyclic graph (DAG). In this framework, each node corresponds to a random variable, while directed edges between nodes signify probabilistic relationships and dependencies.

The joint probability distribution over all variables in the network can be factorized using the chain rule of probability, expressed as:

$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{Parents}(X_i))$

where X_i represents the random variables in the network and $\text{Parents}(X_i)$ denotes the set of parent nodes of X_i within the graph structure.

For illustration, consider a student performance model with variables: Intelligence (I), Difficulty (D), Grade (G), SAT score (S), and Letter of recommendation (L). The joint probability distribution would decompose as:

$P(I, D, G, S, L) = P(I) \cdot P(D) \cdot P(G \mid I, D) \cdot P(S \mid I) \cdot P(L \mid G)$

This factorization demonstrates how the network structure captures conditional independence relationships, enabling efficient computation of posterior probabilities through Bayesian inference while maintaining interpretability of the underlying probabilistic relationships.

Decision Stump

A Decision Stump is a simplified machine learning model for classification tasks, consisting of a single-level decision tree that generates predictions based on a solitary feature threshold. This model partitions the feature space using an optimal split point to separate classes with minimal complexity.

The decision function is formally defined as:

$$h(x) = \begin{cases} 1 & \text{if } x_j \leq \theta \\ 0 & \text{if } x_j > \theta \end{cases}$$

where $h(x)$ represents the predicted class label, x_j denotes the value of the selected feature, and θ is the optimal threshold value determined during training.

Model training involves minimizing the classification error, expressed as:

$E(\theta) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(h(x_i) \models y_i)$

where N represents the total number of training samples, and $\mathbb{I}$ is the indicator function that returns 1 when the prediction $h(x_i)$ mismatches the true label y_i, and 0 otherwise. The optimal threshold θ is selected to minimize this empirical error, establishing the decision stump as an efficient and highly interpretable baseline model for binary classification problems.

Random Forest

Random Forest is an ensemble learning method that constructs multiple decision trees during training and aggregates their predictions through majority voting for classification tasks or averaging for regression problems. This approach enhances predictive accuracy and controls overfitting by introducing randomness through bootstrap sampling of training instances and feature subset selection at each split.

The ensemble prediction $\hat{y}$ for an input x is given by:

$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x)$

where T represents the total number of trees in the forest, and $h_t(x)$ denotes the prediction from the t-th decision tree. Each tree is trained on a bootstrap sample of the original dataset with feature randomization at split points, creating model diversity that improves generalization performance.

Feature importance quantification is a key advantage of Random Forest, calculated as the normalized total reduction in impurity achieved by each feature across all trees:

$$\text{Importance}(i) = \sum_{j \in \text{nodes}} P(j) \cdot \Delta I(j,i)$$

where $P(j)$ represents the proportion of samples reaching node j, and $\Delta I(j,i)$ denotes the impurity decrease (measured by Gini impurity or entropy) attributable to feature i at node j. This ensemble methodology provides robust performance while maintaining interpretability through feature importance rankings.

AdaBoost.M1

AdaBoost.M1 (Adaptive Boosting) is an ensemble learning algorithm that sequentially combines multiple weak classifiers to form a strong classifier for binary classification tasks. The algorithm iteratively trains weak learners, with each subsequent model focusing more on previously misclassified instances through an adaptive weighting mechanism.

The final ensemble prediction $G(x)$ is determined by weighted majority voting:

$$G(x) = \text{sign}\left(\sum_{m=1}^{M} \alpha_m G_m(x)\right)$$

Initialization begins with uniform instance weights $w_i = \frac{1}{N}$ for N training samples. At each iteration m, a weak classifier $G_m(x)$ is trained on the weighted data, and its performance is evaluated through the weighted error rate:

$$\text{Err}_m = \frac{\sum_{i=1}^{N} w_i \mathbb{I}(y^{(i)} \neq G_m(x^{(i)}))}{\sum_{i=1}^{N} w_i}$$

where $\mathbb{I}$ represents the indicator function. The classifier's influence in the final ensemble is determined by its coefficient:

$$\alpha_m = L \cdot \log\left(\frac{1-\text{Err}_m}{\text{Err}_m}\right)$$

with L denoting a learning rate parameter that controls update aggressiveness. Instance weights are then updated to emphasize misclassified samples:

$$w_i \leftarrow \frac{w_i \cdot \exp\left(\alpha_m \cdot \mathbb{I}(y^{(i)} \neq G_m(x^{(i)}))\right)}{\sum_{j=1}^{N} w_j}$$

This iterative process continues for M rounds, progressively refining the model's focus on challenging instances and resulting in a robust classifier that effectively minimizes classification error through adaptive learning.

Instance-Based K-Nearest Neighbors (IBk)

Instance-Based K-Nearest Neighbors (IBk) is an instance-based supervised learning algorithm employed for both classification and regression tasks. The method operates on the principle of similarity, where predictions for new query points are derived from the characteristics of their k closest neighbors in the feature space.

The IBk prediction function is formally defined as:

$$\hat{y} = \begin{cases} \text{mode}(\{y_i\}_{i \in N_k(x)}) & \text{for classification} \\ \frac{1}{k}\sum_{i \in N_k(x)} y_i & \text{for regression} \end{cases}$$

where $N_k(x)$ denotes the set of indices corresponding to the k nearest neighbors of query point x within the training set, and y_i represents the target variable of the i-th neighbor.

Distance computation between instances typically employs metrics such as Euclidean distance for continuous features or Manhattan distance for robust measurement in high-dimensional spaces. While IBk effectively captures complex decision boundaries without parametric assumptions, its computational requirements scale with dataset size, and performance may degrade in high-dimensional settings due to the curse of dimensionality.

Experimental Implementation

All machine learning algorithms were implemented using the Weka toolkit version 3.8.5. Model training and evaluation followed a standardized protocol with a 90% training and 10% testing data split ratio to ensure consistent performance assessment and facilitate optimal model selection across all algorithmic approaches.

Results and Discussion

This section presents a comprehensive evaluation of six machine learning models applied to the Netflix dataset for exploratory data analysis. The study examines the performance of Decision Table (DT), Bayes Net (BN), Decision Stump (DS), Random Forest (RF), AdaBoost M1 (Ada), and IBk (Instance-Based K-Nearest Neighbors) algorithms to identify the most effective approach for the classification task.

Table 2: Performance of Selected Models

No.	Classifiers	Accuracy	Precision	Recall	ROC	PRC	Kappa	F-Measure	MCC	Time
1	DT	88.12%	0.90	0.87	0.88	0.88	0.76	0.87	0.78	0.09
2	BN	88.57%	0.91	0.89	0.95	0.94	0.77	0.89	0.79	0.09
3	DS	94%	0.71	0.70	0.64	0.69	0.40	0.70	0.41	0.99
4	RF	90%	0.92	0.90	0.98	0.97	0.80	0.90	0.82	0.04
5	Ada	97.14%	0.97	0.97	0.96	0.94	0.94	0.97	0.94	0.03
6	IBk	88.09%	0.89	0.88	0.92	0.89	0.76	0.88	0.77	0.02

As detailed in Table 2, the models demonstrated varying levels of performance across multiple evaluation metrics. AdaBoost M1 emerged as the top-performing algorithm, achieving exceptional results with 97.14% accuracy, 0.97 precision, 0.97 recall, 0.96 ROC, and 0.94 PRC. In contrast, IBk showed the lowest accuracy at 88.09%, though it maintained respectable performance with 0.89 precision, 0.88 recall, 0.92 ROC, and 0.89 PRC.

The intermediate performers included Decision Stump with 94% accuracy, Random Forest with 90% accuracy, Bayes Net with 88.57% accuracy, and Decision Table with 88.12% accuracy. Notably, Random Forest achieved the highest ROC (0.98) and PRC (0.97) values among all models, indicating superior discriminatory capability despite its moderate accuracy.

The analysis also considered computational requirements and additional statistical metrics. AdaBoost M1 demonstrated both high performance and efficiency, requiring only 0.03 seconds for model creation while achieving outstanding statistical measures (0.94 kappa, 0.97 F-Measure, 0.94 MCC). Decision Table showed the fastest training time at 0.09 seconds with solid performance metrics (0.76 kappa, 0.87 F-Measure, 0.79 MCC).

Random Forest balanced moderate training time (0.04 seconds) with strong statistical performance (0.80 kappa, 0.90 F-Measure, 0.94 MCC), while IBk achieved the fastest model creation time (0.02 seconds) with competitive metrics (0.76 kappa, 0.78 MCC). Decision Stump, despite its high accuracy, showed relatively weaker statistical consistency (0.40 kappa, 0.70 F-Measure, 0.41 MCC), suggesting potential limitations in model robustness.

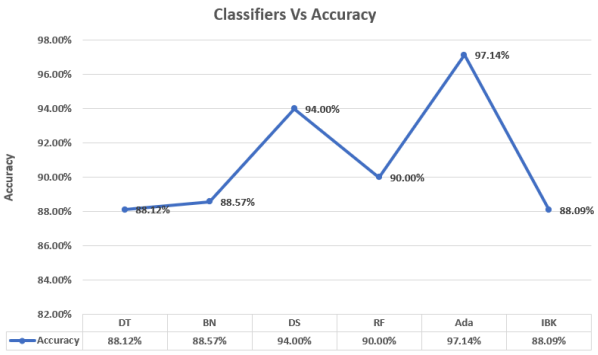

Fig. 2: Model Accuracy Comparison

Figure 2 illustrates the accuracy performance across all models, clearly demonstrating AdaBoost M1's superior classification capability (97.14%) and IBk's position as the least accurate model (88.09%). The performance gradient shows Decision Stump (94%) significantly outperforming Random Forest (90%), with Bayes Net (88.57%) and Decision Table (88.12%) clustering in the lower accuracy range. This comparative analysis provides valuable insights for model selection based on the specific requirements of accuracy, computational efficiency, and statistical reliability.

Figure 3 illustrates the precision performance across all evaluated models. AdaBoost demonstrated superior precision (0.97), effectively minimizing false positive predictions, while Decision Stump recorded the lowest precision (0.71). Intermediate precision values were observed for Random Forest (0.92), Bayes Net (0.91), Decision Table (0.90), and IBk (0.89), indicating varying capabilities in positive class identification accuracy across different algorithmic approaches.

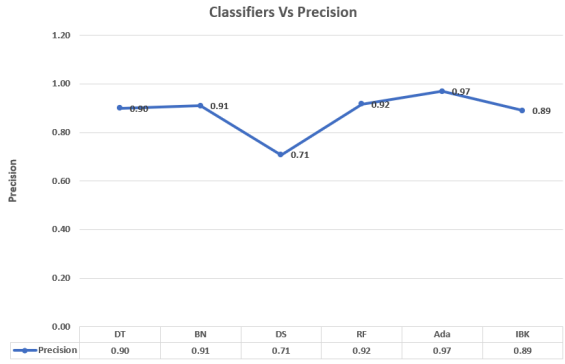

Fig. 3: Model Precision Comparison

As shown in Figure 4, recall metrics reveal similar performance patterns, with AdaBoost achieving the highest recall (0.97), demonstrating exceptional sensitivity in identifying true positive instances. Decision Stump again showed the weakest performance (0.70 recall), while Random Forest (0.92), Bayes Net (0.91), and Decision Table (0.88) displayed moderate to strong recall capabilities, balancing sensitivity with specificity across the classification task.

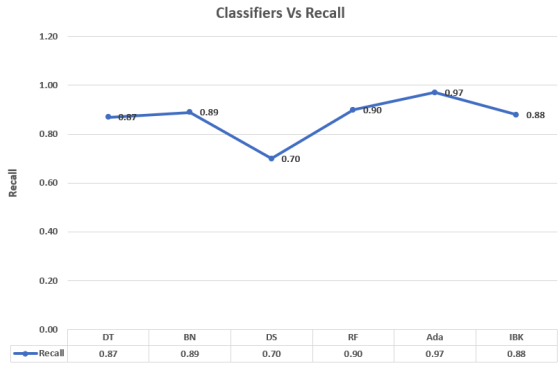

Fig. 4: Model Recall Comparison

Figures 5 and 6 present receiver operating characteristic (ROC) and precision-recall curve (PRC) metrics, respectively. Random Forest achieved the highest scores in both categories (ROC: 0.98, PRC: 0.97), indicating superior overall discriminatory power and precision-recall balance. In contrast, Decision Stump recorded the lowest values (ROC: 0.64, PRC: 0.69), reflecting limited classification capability. AdaBoost maintained strong performance in both metrics (ROC:

DOI: 10.3844/ojbsci.2025.2191.2203

0.96, PRC: 0.94), followed by Bayes Net (ROC: 0.95, PRC: 0.94), IBk (ROC: 0.92, PRC: 0.89), and Decision Table (ROC: 0.88, PRC: 0.88).

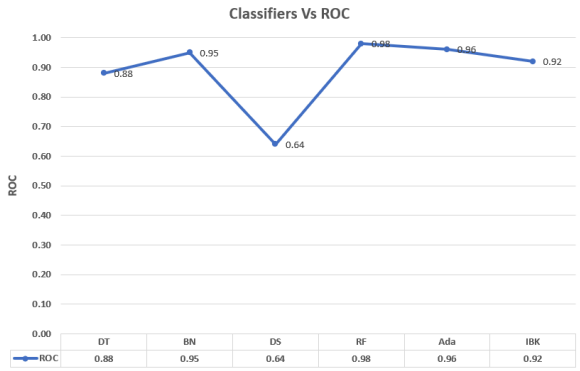

Fig. 5: Model ROC Comparison

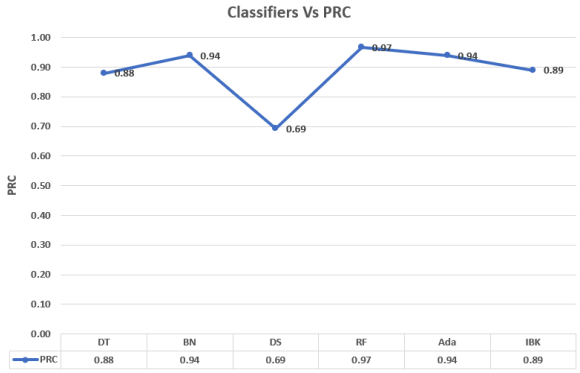

Fig. 6: Model PRC Comparison

Figure 7 displays Cohen's Kappa statistics, measuring agreement between predicted and actual classifications beyond chance. AdaBoost again demonstrated exceptional performance (0.94), indicating nearly perfect agreement, while Decision Stump showed only fair agreement (0.40). Random Forest (0.80), Bayes Net (0.77), Decision Table (0.76), and IBk (0.76) exhibited substantial agreement levels, confirming the general reliability of their classification patterns across the dataset.

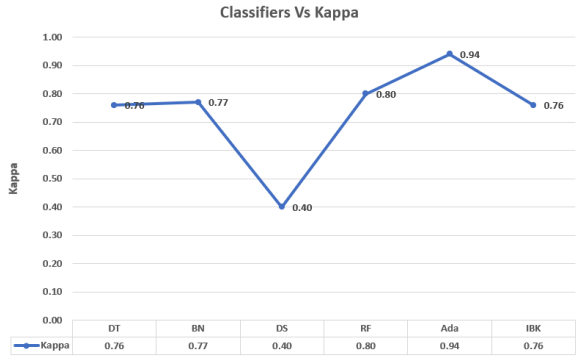

Fig. 7: Model Kappa Comparison

Figure 8 presents the F-Measure results, demonstrating the harmonic mean between precision and recall across all models. AdaBoost achieved the highest F-Score (0.97), indicating exceptional balance between precision and recall, while Decision Stump recorded the lowest (0.70). The remaining models showed moderate performance: Random Forest (0.90), Bayes Net (0.89), IBk (0.88), and Decision Table (0.87).

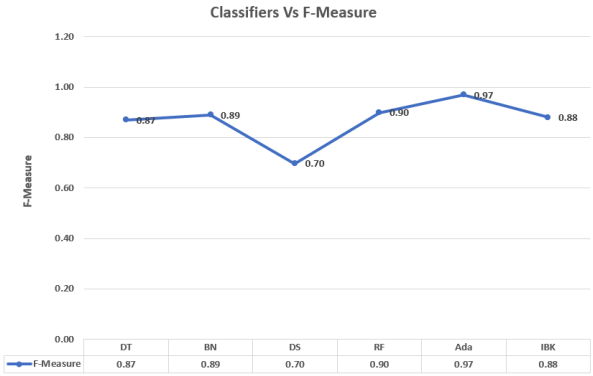

Fig. 8: Model F-Measure Comparison

Figure 9 displays Matthews Correlation Coefficient (MCC) values, which provide a more reliable statistical measure for binary classification. AdaBoost again demonstrated superior performance (0.94 MCC), followed by Random Forest (0.82), Bayes Net (0.79), Decision Table (0.78), IBk (0.77), with Decision Stump showing the weakest correlation (0.41).

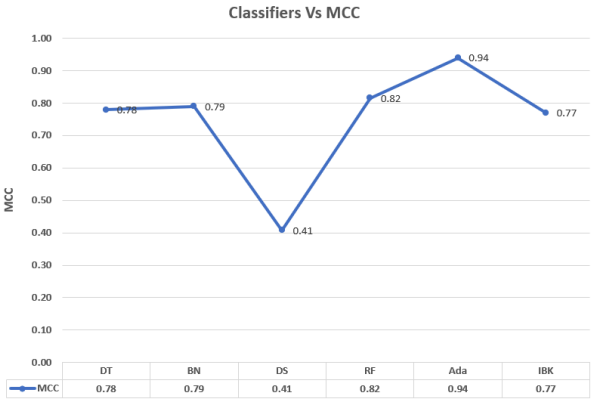

Fig. 9: Model MCC Comparison

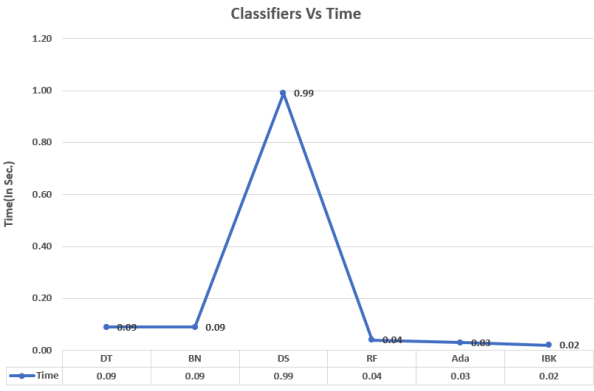

Fig. 10: Classifiers vs Time Consumption

DOI: 10.3844/ojbsci.2025.2191.2203

Figure 10 illustrates the time consumption required for model training and evaluation. The analysis reveals significant variation in computational demands across algorithms, with ensemble methods demonstrating efficient processing times relative to their performance gains.

Table 3 presents comprehensive error metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE). AdaBoost consistently outperformed other models across all error metrics, achieving the lowest MAE (0.05), RMSE (0.17), RAE (9.84%), and RRSE (33.39%).

Decision Stump exhibited the highest error rates with MAE (0.37), RMSE (0.47), RAE (74.74%), and RRSE (94.88%). The remaining models demonstrated moderate error performance: Random Forest (MAE: 0.10, RMSE: 0.29), Decision Table (MAE: 0.10, RMSE: 0.23), Bayes Net (MAE: 0.11, RMSE: 0.25), and IBk (MAE: 0.12, RMSE: 0.32).

Table 3: Deviations of probabilistic learning with time complexity

No.	Classifiers	MAE	RMSE	RAE	RRSE
1	DT	0.10	0.23	21.61%	48.20%
2	BN	0.11	0.25	21.68%	49.20%
3	DS	0.37	0.47	74.74%	94.88%
4	RF	0.10	0.29	20%	57.27%
5	Ada	0.05	0.17	9.84%	33.39%
6	IBk	0.12	0.32	23.64%	64.58%

Figure 11 specifically highlights Mean Absolute Error performance, reinforcing AdaBoost's superior predictive accuracy (0.05 MAE) and Decision Stump's limitations (0.37 MAE). The clustering of Decision Table (0.10), Random Forest (0.10), Bayes Net (0.11), and IBk (0.12) within a narrow MAE range suggests comparable baseline error rates for these algorithms.

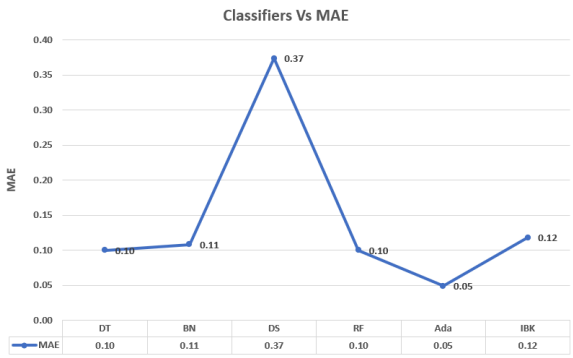

Fig. 11: Classifiers vs MAE

Figure 12 presents the Root Mean Squared Error performance across all evaluated models. AdaBoost demonstrated superior predictive accuracy with the lowest RMSE (0.17), while Decision Stump recorded the highest error (0.47 RMSE). The remaining models showed intermediate RMSE values: Decision Table (0.23), Bayes Net (0.25), Random Forest (0.29), and IBk (0.32), indicating varying levels of prediction variance across different algorithmic approaches.

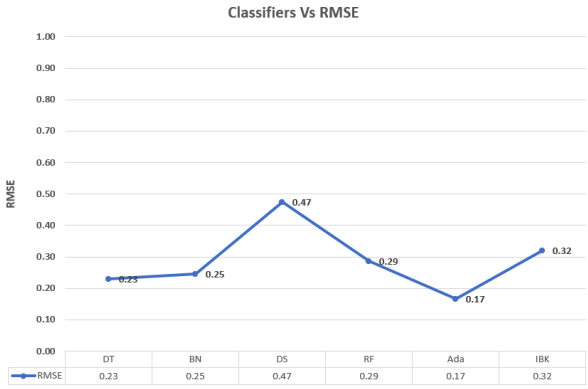

Fig. 12: Classifiers vs RMSE

As illustrated in Figure 13, Relative Absolute Error metrics reveal similar performance patterns. AdaBoost again achieved the lowest RAE (9.84%), significantly outperforming other models. Decision Stump exhibited the highest relative error (74.74%), while the remaining models clustered within a moderate range: Random Forest (20.00%), Decision Table (21.61%), Bayes Net (21.68%), and IBk (23.64%).

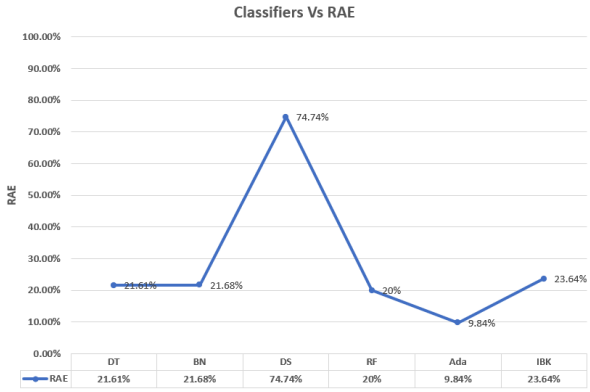

Fig. 13: Classifiers vs RAE

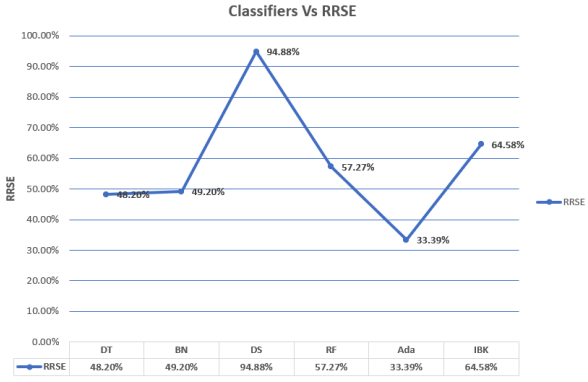

Fig. 14: Classifiers vs RRSE

Figure 14 displays Root Relative Squared Error performance, completing the comprehensive error analysis. AdaBoost maintained its superior performance

with the lowest RRSE (33.39%), followed by Random Forest (57.27%), Decision Table (48.20%), Bayes Net (49.20%), and IBk (64.58%). Decision Stump again demonstrated the weakest performance with the highest RRSE (94.88%), reinforcing its limitations for this classification task.

Conclusion

This comprehensive study evaluated six machine learning classifiers, Decision Table (DT), Bayesian Network (BN), Decision Stump (DS), Random Forest (RF), AdaBoost (Ada), and K-Nearest Neighbors (IBk), across multiple performance dimensions, revealing significant variations in algorithmic effectiveness for the Netflix dataset analysis.

AdaBoost demonstrated exceptional performance across nearly all evaluation metrics, achieving the highest accuracy (97.14%), precision (0.97), recall (0.97), F-measure (0.97), and MCC (0.94), along with the lowest error rates in regression tasks (MAE: 0.05, RMSE: 0.17). This consistent superiority establishes AdaBoost as the optimal choice for applications requiring maximum predictive accuracy.

Random Forest emerged as a strong alternative, balancing high performance (accuracy: 90%, MCC: 0.82) with exceptional computational efficiency (0.04 seconds training time). Bayesian Network distinguished itself with superior probabilistic discrimination capabilities, achieving the highest ROC (0.95) and PRC (0.94) scores. Conversely, Decision Stump illustrated the performance trade-offs of simplified models, delivering rapid training (0.99 seconds) but substantially lower predictive accuracy across all metrics.

While this study demonstrates the effectiveness of traditional machine learning approaches for structured data analysis, it acknowledges the growing potential of deep learning architectures, including RNNs, LSTMs, CNNs, and Transformer models, for handling complex sequential and textual data. The absence of these advanced techniques presents an opportunity for future comparative research.

The findings provide practical guidance for model selection based on specific application requirements, whether prioritizing accuracy (AdaBoost), speed (Random Forest), or probabilistic calibration (Bayesian Network). Future research should explore the impact of advanced feature engineering, automated hyperparameter optimization, and hybrid ensemble methods. Additionally, validating these models on larger, more diverse datasets would further establish their generalization capabilities and real-world applicability.

This comparative analysis contributes to the growing body of knowledge on machine learning applications in media analytics, offering empirical evidence to support informed algorithmic selection for content recommendation and classification tasks in streaming platforms.

Funding

The authors did not receive any funding for this study.

Author's Contributions

All authors equally contributed to this study.

Ethics

This manuscript is an original work. The corresponding author certifies that all co-authors have reviewed and approved the final version of the manuscript. No ethical concerns are associated with this article

References

Alfayad, K., Murray, R. L., Britton, J., & Barker, A. B. (2022). Content Analysis of Netflix and Amazon Prime Instant Video Original Films in the Uk for Alcohol, Tobacco and Junk Food Imagery. *Journal of Public Health*, *44*(2), 302–309. https://doi.org/10.1093/pubmed/fdab022

Allem, J.-P., Van Valkenburgh, S. P., Donaldson, S. I., Dormanesh, A., Kelley, T. C., & Rosenthal, E. L. (2022). E-Cigarette Imagery in Netflix Scripted Television and Movies Popular Among Young Adults: A Content Analysis. *Addictive Behaviors Reports*, *16*, 100444. https://doi.org/10.1016/j.abrep.2022.100444

Arti (2021). Netflix or Amazon Prime: who uses AI and machine learning better? Analytics Insight.

Barker, A. B., Smith, J., Hunter, A., Britton, J., & Murray, R. L. (2019). Quantifying Tobacco and Alcohol Imagery in Netflix and Amazon Prime Instant Video Original Programming Accessed from the Uk: A Content Analysis. *BMJ Open*, *9*(2), e025807. https://doi.org/10.1136/bmjopen-2018-025807

Behera, G., & Nain, N. (2022). DeepNNMF: Deep Nonlinear Non-Negative Matrix Factorization to Address Sparsity Problem of Collaborative Recommender System. *International Journal of Information Technology*, *14*(7), 3637–3645. https://doi.org/10.1007/s41870-022-00982-1

Castro, D., Rigby, J. M., Cabral, D., & Nisi, V. (2021). The Binge-Watcher's Journey: Investigating Motivations, Contexts, and Affective States Surrounding Netflix Viewing. *Convergence: The International Journal of Research into New Media Technologies*, *27*(1), 3–20. https://doi.org/10.1177/1354856519890856

Çelik, H. C. (2023). Representations of Older People in Turkish Prime-Time Tv Series and Netflix Original Turkish Series: A Comparative Content Analysis. *Journal of Aging Studies*, *66*, 101158. https://doi.org/10.1016/j.jaging.2023.101158

Frey, M. (2021). *Netflix Recommends: Algorithms, Film Choice, and the History of Taste*(1st ed.). University of California Press. https://doi.org/10.2307/j.ctv269fvqp

Gomez-Uribe, C. A., & Hunt, N. (2016). The Netflix Recommender System: Algorithms, Business Value and Innovation. *ACM Transactions on Management Information Systems (TMIS)*, *6*(4), 1–19. https://doi.org/10.1145/2843948

Hanewinkel, R., Neumann, C., & Morgenstern, M. (2023). Rauchen in Netflix-Spielfilmen und Jugendschutz Smoking in Netflix Feature Films and Youth Protection. *Pneumologie*, *77*(07), 408–412. https://doi.org/10.1055/a-2081-0989

Kavitha, P., Ayyappan, G., Jayagopal, P., Mathivanan, S. K., Mallik, S., Al-Rasheed, A., Alqahtani, M. S., & Soufiene, B. O. (2023). Detection for Melanoma Skin Cancer Through ACCF, BPPF and CLF Techniques with Machine Learning Approach. *BMC Bioinformatics*, *24*(1), 458. https://doi.org/10.1186/s12859-023-05584-7

Khalique, A., Rahmani, M. K. I., Saquib, M., Hussain, I., Muzaffar, A. W., Ahad, Mohd. A., Nafis, M. T., & Ahmad, M. W. (2022). A Deterministic Model for Determining Degree of Friendship Based on Mutual Likings and Recommendations on OTT Platforms. *Computational Intelligence and Neuroscience*, *2022*, 1–11. https://doi.org/10.1155/2022/9576468

Mier, J., & Kohli, A. K. (2021). Netflix: Reinvention Across Multiple Time Periods, Reflections and Directions for Future Research. *AMS Review*, *11*(1–2), 194–205. https://doi.org/10.1007/s13162-021-00197-w

Rastogi, D., Parihar, T. S., & Kumar, H. (2023). A Parametric Analysis of AVA to Optimise Netflix Performance. *International Journal of Information Technology*, *15*(5), 2687–2694. https://doi.org/10.1007/s41870-023-01281-z

Salman, O. H., Taha, Z., Alsabah, M. Q., Hussein, Y. S., Mohammed, A. S., & Aal-Nouman, M. (2021). A Review on Utilizing Machine Learning Technology in the Fields of Electronic Emergency Triage and Patient Priority Systems in Telemedicine: Coherent Taxonomy, Motivations, Open Research Challenges and Recommendations for Intelligent Future Work. *Computer Methods and Programs in Biomedicine*, *209*, 106357. https://doi.org/10.1016/j.cmpb.2021.106357