Poly-BoostNet Enhanced Consumer Churn Analytics With CART-Driven Insights for Business Decision Systems

D. Ramireddy[1], Gurram Murali[2], Chevva Anjali[2], Umme Aiman Faiza[2], Avula Karthik[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (Data science)

[1,2]Vaagdevi Engineering College, Bollikunta, Warangal, 506005, Telangana, India

To Cite this Article

D. Ramireddy, Gurram Murali, Chevva Anjali, Umme Aiman Faiza, Avula Karthik, "Poly-BoostNet Enhanced Consumer Churn Analytics With CART-Driven Insights for Business Decision Systems", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 04, April 2026,pp: 398-409, DOI: http://doi.org/10.64771/jsetms.2026.v03.i04.pp398-409

Submitted: 28-02-2026 Accepted: 01-04-2026 Published: 09-04-2026

ABSTRACT

Customer churn prediction has emerged as a vital analytical process for organizations seeking to improve customer retention, strengthen marketing strategies, and ensure stable revenue growth in competitive markets. With the rapid increase in digital customer interactions, there is a growing need for intelligent systems that can accurately identify customers who are likely to disengage or reduce their activity. Conventional churn analysis methods relied on manual evaluation using simple statistical techniques such as averages, rule-based thresholds, and spreadsheet analysis, which are insufficient for capturing complex behavioral patterns in large datasets. To overcome these challenges, this study introduces a machine learning-based predictive framework built on a Classification and Regression Tree (CART) approach. The system utilizes supervised learning to analyze multiple customer-related features, including demographic information, purchase history, and digital engagement behavior, to predict both churn risk and customer satisfaction levels. The framework evaluates several models, including Passive-Aggressive (PA), Support Vector Machine (SVM), Extra Trees (ET), and a novel hybrid model called Poly-BoostNet. The Poly-BoostNet model integrates Recurrent Polynomial Network (RPN) feature expansion with Categorical Boosting (CB), allowing it to effectively capture nonlinear relationships and complex feature interactions. Experimental results indicate that while PA and SVM offer faster learning and moderate accuracy, Poly-BoostNet achieves superior performance, delivering the highest R^2 score in regression tasks and achieving 100% accuracy in classification of customer engagement. The system is developed using Python-based machine learning techniques and deployed via a Flask web application, providing a complete pipeline for data preprocessing, model training, evaluation, and real-time prediction. This framework offers a scalable and efficient solution for data-driven customer retention strategies.

Keywords: Customer Churn Prediction, Machine Learning, Classification and Regression Tree (CART), Supervised Learning, Poly-BoostNet, Recurrent Polynomial Network (RPN), Categorical Boosting (CB).

1.INTRODUCTION

Customer retention is a significant challenge for organizations across multiple industries such as telecommunications, retail, banking, insurance, healthcare, education, and subscription-based services. A major issue contributing to this challenge is customer churn [1], which occurs when customers terminate their relationship with a company. High churn rates can lead to substantial revenue loss, with certain sectors reporting annual churn levels between 20% and 40% [2], as shown in fig 1 Furthermore, studies reveal that the cost of acquiring new customers is considerably higher often five to twenty-five times than retaining existing ones, making churn reduction a critical business objective [3].

Fig. 1: Customer Retention Strategies

To address this issue, Machine Learning and Deep Learning techniques have become increasingly important in churn prediction. These approaches are capable of handling large-scale, complex, and high-dimensional customer data [4], enabling more accurate identification of patterns and trends. Traditional churn prediction methods, including rule-based approaches and basic statistical models, are often limited in their ability to capture the complexity of customer behaviour [5]. In contrast, advanced learning algorithms can model intricate relationships within the data [6], making them more effective for developing accurate and scalable churn prediction systems.

2.LITERATURE SURVEY

Imani et al. [7] researched the application of Machine Learning (ML) and Deep Learning (DL) techniques for churn prediction by reviewing 240 relevant studies, with 61 selected for detailed qualitative synthesis. They found that while ensemble methods like XGBoost remained dominant in ML, DL approaches such as LSTM and CNN were increasingly applied to complex datasets. However, the study highlighted that issues like class imbalance, interpretability challenges, and limited use of profit-oriented metrics persisted, despite the potential shown by Explainable AI and adaptive learning methods. Ajegbile et al. [8] analysed CRM for various applications, helping businesses proactively retain at-risk customers and maximize customer lifetime value. With high churn rates leading to substantial revenue losses, businesses in subscription-based services, telecommunications, retail, banking, education, healthcare, Insurance, and other sectors increasingly rely on data-driven approaches to enhance customer retention strategies.

Janssens et al. [9] introduced B2Boost, an instance-dependent gradient boosting model explicitly designed for B2B churn scenarios. Recognizing customer heterogeneity in profitability, they propose the Expected Maximum Profit for B2B churn (EMPB) metric to guide model training. B2Boost directly optimizes customer-specific profit rather than traditional classification accuracy, yielding notable profit improvements over standard approaches. Shimaa et al. [10] developed a hybrid churn prediction framework that combines XGBoost with SMOTE-ENN resampling to balance datasets and improve classification accuracy. This integration enhances precision, recall, and F1 scores, outperforming conventional ML techniques across three telecom datasets. By effectively addressing class imbalance and leveraging ensemble learning, the model facilitates proactive retention strategies, reinforcing the role of resampling techniques in churn prediction.

Lee et al. [11] proposed a hybrid churn prediction framework that dynamically models churn probability based on customer lifetime value rather than fixed periods. By segmenting customers into groups such as new, short-term, high-value, and churn-prone users, their methodology applies tailored ML models to enhance predictive accuracy. Evaluations of datasets from a U.K. gift seller and Pakistan's most significant e-commerce platform show recall scores ranging from 0.56 to 0.72 in one case and 0.91 to 0.95 in another. The study highlights the advantages of integrating statistical modelling with ML techniques to refine customer retention strategies while reducing data requirements. Wang et al. [12] addressed the challenge of player churn prediction in online video games, where understanding social interaction dynamics is critical. While ML models are widely used for player behaviour analysis, their black-box nature limits adoption by product managers and game designers. The study restructures model inputs into explicit and implicit features to bridge this gap, enhancing expert interpretability. Furthermore, the research highlights the necessity of XAI techniques that explain feature contributions and provide actionable recommendations for reducing churn.

Babak et al. [13] introduced a social network-based churn prediction model, recognizing that social interactions and peer behaviour often influence customer churn. The study develops a feature engineering approach incorporating influence and conformity indices derived from call network data. By integrating social connectivity metrics, the model significantly enhances the predictive power of standard ML classifiers, particularly gradient boosting models. This research demonstrates that churn is not solely an individual decision but is shaped by broader social dynamics. This perspective extends beyond telecommunications to various industries were peer influence affects customer behaviour. Šimović et al. [14] explored churn prediction using big data analytics to analyze heterogeneous customer behaviours, such as self-care service usage, service duration, and responsiveness to marketing efforts. Their study introduces an enhanced logistic regression model with a mixed penalty term to mitigate overfitting and balance feature selection. Empirical evaluation on a large CRM dataset demonstrates high classification performance across standard metrics, reinforcing the potential of penalized logistic regression as a scalable and computationally efficient approach to churn modeling in big data environments.

Jakob et al. [15] extended traditional ML techniques to the digital health sector, investigating early user churn in a weight loss app. By analyzing engagement data from 1283 users and 310,845 event logs, the study employs an RF model to predict user dropout based on daily login counts. Achieving an F1 score of 0.87 on day 7 and identifying 93% of churned users, the study highlights how churn prediction can enable personalized retention strategies in digital health interventions, ultimately improving long-term user engagement and health outcomes.

3. PROPOSED SYSTEM

The proposed methodology establishes a comprehensive, data-driven framework for analyzing customer behavior, specifically targeting advertisement engagement and satisfaction levels. The architecture follows a systematic pipeline that integrates a web-based user interface with a robust Flask-driven analytical engine. By combining traditional machine learning benchmarks with a high-performance hybrid Poly-BoostNet model which fuses RPN for nonlinear feature expansion with CB gradient boosting the system ensures high predictive accuracy across both classification and regression tasks, as shown in fig 2. The workflow encompasses seamless data ingestion, standardized preprocessing, and an adaptive retraining mechanism to ensure long-term model reliability.

User Interface (Web Browser)

- The user interacts with the system through a browser-based graphical interface built using HTML templates.
- Users can perform a wide range of operations including registration, login, data input, classification, regression, and model comparison.
- The interface provides real-time visualization of EDA plots, confusion matrices, ROC curves, and regression performance graphs.
- All user actions are converted into HTTP requests and sent to the Flask backend server for processing.

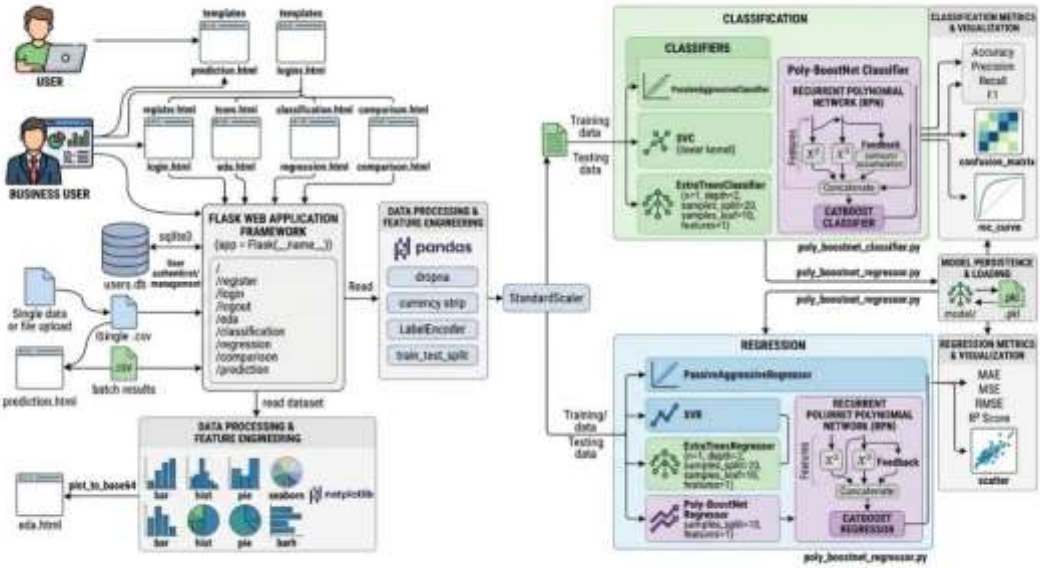

Fig. 2: System architecture.

Flask Web Server (app.py)

- The Flask backend acts as the core controller of the system, handling all incoming user requests and routing.
- It manages user authentication, session tracking, and the complex interaction between different software modules.
- The server coordinates communication between the user interface, machine learning models, and the database.
- It is responsible for model training, loading saved weights, generating predictions, and rendering results back to the frontend.

SQLite Database (users.db)

- The database stores persistent user information such as registration details and login credentials.

- Passwords are securely stored using hashing techniques (such as PBKDF2 or SHA-256) to ensure data protection.
- The Flask server interacts with the database for authentication, user validation, and secure session management.
- It provides a lightweight and efficient storage mechanism for standard system operations.

Raw Data (CSV Input)

- The dataset acts as the primary input source for analysis and prediction.
- It contains critical customer-related attributes including demographics, purchase behavior, engagement with advertisements, and satisfaction levels.
- The dataset is loaded from CSV files and passed directly into the preprocessing pipeline to form the foundation for both classification and regression tasks.

Data Preprocessing & Feature Engineering

- The dataset undergoes cleaning by removing missing values and inconsistent entries to ensure data quality.
- Categorical features are converted into numerical form using Label Encoding techniques.
- Numerical features are standardized using scaling methods (such as MinMaxScaler or StandardScaler) to improve model convergence.
- Final feature vectors are generated and passed to the machine learning models for training and prediction.

Existing Baseline Models

The processed dataset is input into multiple baseline machine learning models to establish performance benchmarks:

- **Passive-Aggressive (PA):** Provides fast online learning for classification.
- **Support Vector Machine (SVM/SVR):** Offers stable decision boundaries for both classification and regression.
- **Extra Trees (ET):** An ensemble-based model that improves robustness and reduces the risk of overfitting.
- These models generate predictions for Customer Engagement with Advertisements (Classification) and Customer Satisfaction Score (Regression).

Proposed Hybrid Poly-BoostNet Model

This is the core intelligent model of the system, combining advanced feature expansion with gradient boosting:

1. **RPN:** Performs polynomial feature expansion to capture complex nonlinear relationships.
    - Applies a cumulative feedback mechanism to simulate sequential dependencies within the data.

2. **CB Learning:** Uses gradient boosting on decision trees for high-accuracy prediction.
    - Handles categorical and numerical data efficiently while reducing overfitting through symmetric tree structures.

Prediction Results & Target Output

- The system generates dual predictions:
    - **Target 1:** Engagement with Advertisements (Classification).
    - **Target 2:** Customer Satisfaction (Regression).
- Results are displayed on the user interface along with detailed performance metrics and model-wise comparisons for analytical depth.

Model Comparison & Evaluation

- The system evaluates all models using standardized metrics:
    - **Classification:** Accuracy, Precision, Recall, and F1-score.
    - **Regression:** MAE, MSE, RMSE, and R2 Score.
- Comparative analysis identifies the best-performing model, typically demonstrating the superior performance of Poly-BoostNet over baseline models.

Model Retraining Mechanism

- The system supports an adaptive retraining cycle using updated customer datasets.
- New data is passed through the same preprocessing and training pipeline, allowing models to adapt to changing consumer trends.
- Updated models are saved and reused for future predictions, ensuring the system remains relevant and accurate over time.

3. RESULTS ANALYSIS

The results analysis section evaluates the performance and effectiveness of the proposed system in achieving accurate and reliable outcomes. It focuses on assessing the model using various evaluation metrics such as accuracy, precision, recall, and F1-score to ensure comprehensive performance measurement. The analysis also compares the proposed approach with existing methods to highlight improvements and advantages. Graphical representations and visualizations are utilized to clearly interpret the results and identify patterns or trends. Additionally, the robustness and generalization capability of the model are examined using test datasets. This section provides critical insights into the strengths and limitations of the system, ensuring its suitability for real-world applications.

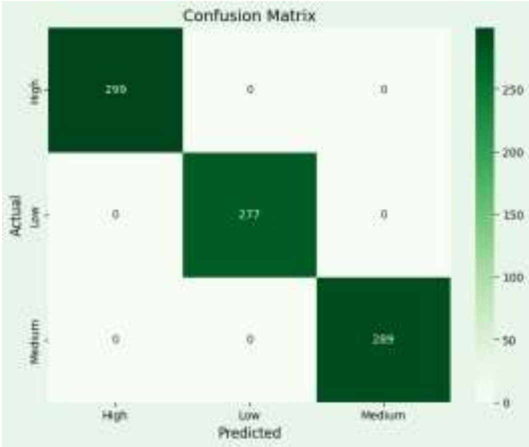

Fig. 3: Confusion Matrices of Various Classifiers - Poly-BoostNet

Fig 3 displays the confusion matrix for the proposed Poly-BoostNet model, which demonstrates significantly improved classification performance. The matrix shows 299 High engagement samples correctly classified, 277 Low engagement samples correctly identified, and 289 Medium engagement samples correctly predicted, with no misclassification across categories. All off-diagonal elements are zero, indicating perfect separation between classes. This result demonstrates that the Poly-BoostNet model effectively captures complex feature relationships using RPN feature extraction combined with CB, leading to highly accurate classification of advertisement engagement levels.

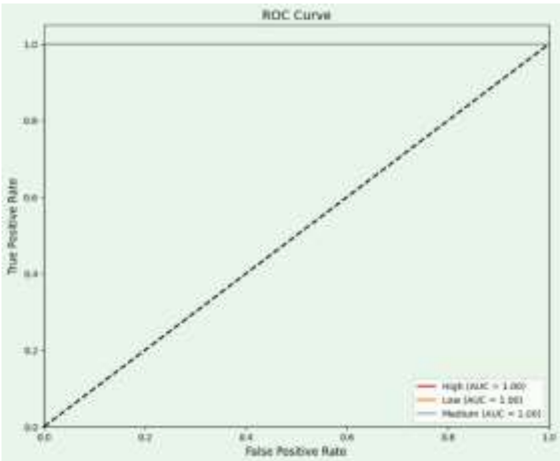

Fig. 4: RoC Curves of Various Classifiers - Poly-BoostNet.

Fig 4 shows the ROC curves for the proposed Poly-BoostNet model, which integrates RPN feature extraction with CB. The ROC curves for all three engagement classes High, Low, and Medium achieve an AUC value of 1.00, indicating perfect classification performance. The curves reach the top-left corner of the ROC space, representing a True Positive Rate of 1 with a False Positive Rate of 0. This result demonstrates that the Poly-BoostNet model successfully learns complex feature interactions and produces highly accurate predictions for advertisement engagement levels compared to the existing classifiers.

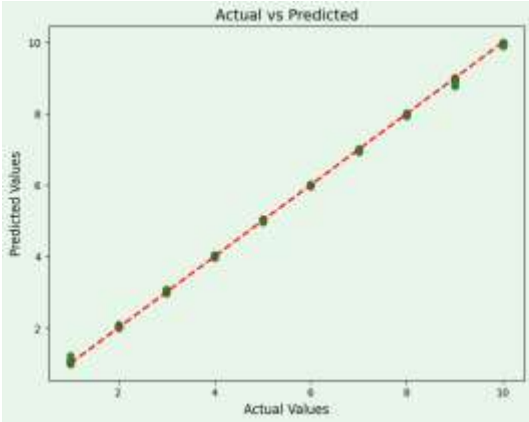

Fig. 5: Scatter Plots of Various Regressors - Poly-BoostNet

Fig.5 displays the scatter plot for the proposed Poly-BoostNet regressor, which combines Recurrent Polynomial Network feature extraction with CB regression. The predicted values align almost perfectly with the diagonal reference line, indicating that the predicted satisfaction scores match the actual values very closely. Data points are distributed directly along the line from 1 to 10, demonstrating minimal prediction error across the entire range of satisfaction scores. This strong alignment confirms that the Poly-BoostNet model successfully captures complex feature relationships and significantly improves regression accuracy compared to the existing PA, SVM, and ET models.

Fig 6 illustrates the single input prediction interface of the developed Poly-BoostNet based consumer churn analytics system. The interface allows users to select both the classification model and regression model, where the Poly-BoostNet model is selected for both tasks. The user provides various feature inputs corresponding to customer behavioral and demographic attributes used by the machine learning models. The entered values include Age = 22, Gender = Female, Income Level = Middle, Marital Status = Married, Education Level = Bachelor’s, Occupation = Middle, Location = Évry, Purchase Category = Gardening & Outdoors, Purchase Amount = $333.80, Frequency of Purchase = 4, Purchase Channel = Mixed, Brand Loyalty = 5, Product Rating = 5, Time Spent on Product Research = 2 hours, Social Media Influence = None, Discount Sensitivity = Somewhat Sensitive, Return Rate = 1, Device Used for Shopping = Tablet, Payment Method = Credit Card, Time of Purchase = 03-01-2024, Discount Used = True, Customer Loyalty Program Member = False, Purchase Intent = Need-based, Shipping Preference = No Preference, and Time to Decision = 2. These input parameters represent the customer’s purchasing behavior, demographic characteristics, and marketing interaction indicators, which are then processed by the trained models to predict Engagement with Advertisements (classification output) and Customer Satisfaction Score (regression output).

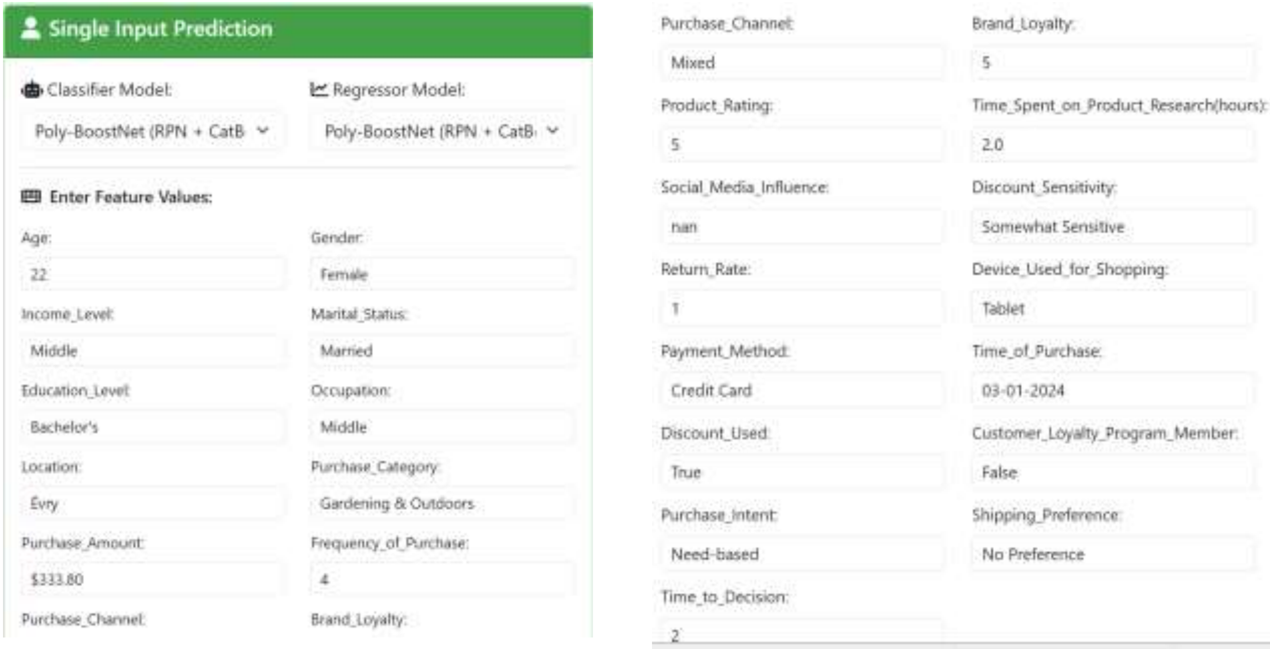

Fig. 6. Prediction input as single data.

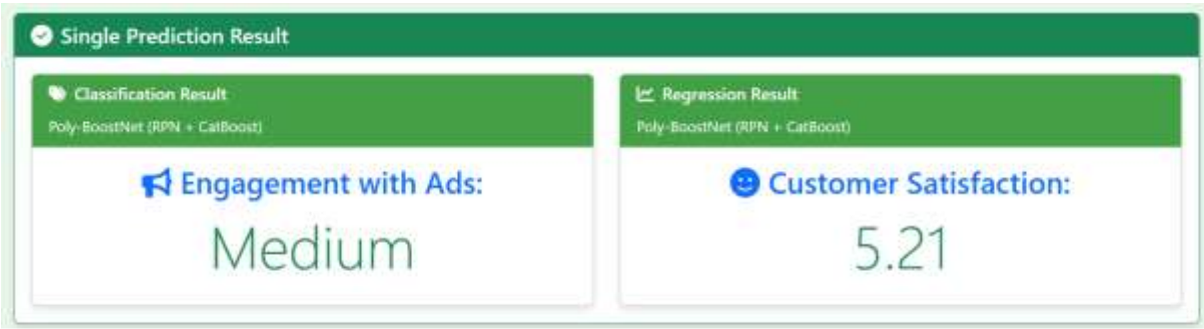

Fig. 7: Predicted outputs from single input.

Fig 7 presents the prediction results generated by the proposed Poly-BoostNet model after processing the single input values provided by the user. The system produces two outputs simultaneously: a classification result for advertisement engagement and a regression result for customer satisfaction. In the classification output, the model predicts the Engagement with Ads level as “Medium”, indicating that the given customer profile is moderately responsive to marketing advertisements. In the regression output, the model predicts a Customer Satisfaction score of 5.21, which falls within the moderate satisfaction range on the scale of 1 to 10. These results demonstrate how the system utilizes the input features related to customer demographics, purchasing behaviour, and marketing interaction patterns to generate meaningful predictions. The interface displays the results clearly using separate panels for classification and regression, enabling business analysts to easily interpret customer engagement behaviour and satisfaction levels for decision-making and targeted marketing strategies.

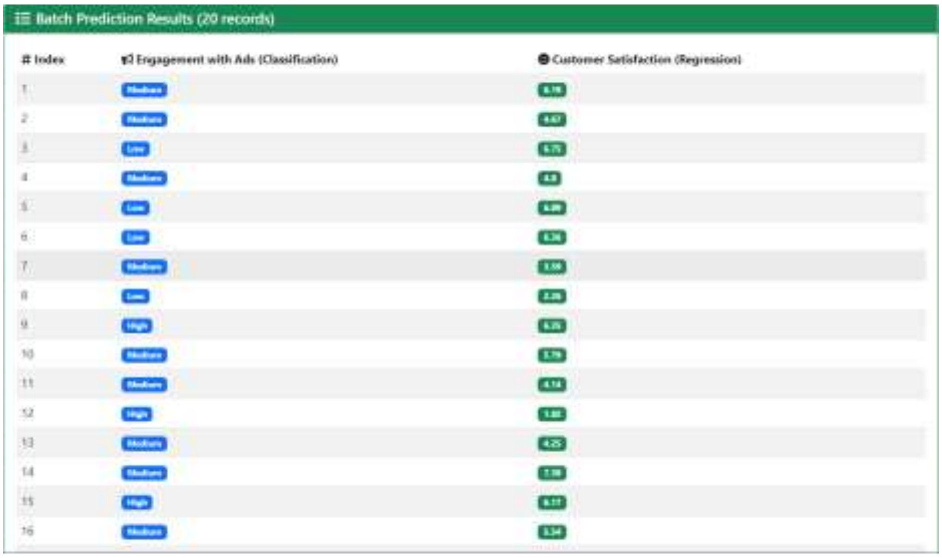

Fig. 8: Batch Prediction Output Results from CSV File.

Fig 8 presents the batch prediction results generated by the Poly-BoostNet model after processing the uploaded CSV file containing multiple customer records. The interface displays the results in a tabular format, where each row corresponds to an individual customer record from the uploaded dataset. The table includes three main columns: Index, Engagement with Ads (Classification), and Customer Satisfaction (Regression). For example, the first record predicts medium engagement with a satisfaction score of 6.19, while the second record shows medium engagement with a satisfaction value of 4.67. Other examples include Low engagement with satisfaction scores of 6.75, 6.89, and 6.36, and High engagement predictions with scores such as 6.35 and 6.17. The predicted satisfaction scores vary across the range of approximately 1.82 to 7.39, indicating different levels of customer experience. This batch prediction functionality enables organizations to analyse multiple customer profiles simultaneously, allowing businesses to quickly identify engagement levels and satisfaction trends across large datasets for effective marketing and decision-making.

Table 1. Classification Models Comparison of Various Methodologies.

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
PA	37.8	37.93	37.8	37.74
SVM	39.19	39.47	39.19	38.82
ET Classifier	34.91	23.53	34.91	26.77
Poly-BoostNet	100.0	100.0	100.0	100.0

Table 1 presents the performance comparison of different classification models used to predict Engagement with Advertisements. The PA classifier achieves an accuracy of 37.8%, with precision 37.93%, recall 37.8%, and F1-score 37.74%, indicating limited predictive capability. The SVM slightly improves performance with an accuracy of 39.19%, precision 39.47%, recall 39.19%, and F1-score 38.82%, demonstrating marginal improvement over the PA model. The ET classifier shows the lowest performance, achieving 34.91% accuracy, 23.53% precision, 34.91% recall, and 26.77% F1-score, suggesting difficulty in correctly distinguishing engagement categories. In contrast, the proposed Poly-BoostNet model significantly outperforms all existing methods, achieving 100%

accuracy, precision, recall, and F1-score, indicating perfect classification of advertisement engagement levels.

5. Conclusion

This research presented an advanced machine learning framework for customer churn analytics and advertisement engagement prediction using a hybrid model called Poly-BoostNet. The system integrates RPN feature extraction with CB and CART-based learning to capture complex relationships within customer behavioural data. The proposed framework was implemented using a Flask-based web application with role-based access for business analysts and end users, enabling functionalities such as exploratory data analysis, classification, regression, model comparison, and real-time prediction. The dataset included various customer attributes such as demographic information, purchasing behaviour, income levels, and marketing engagement indicators, which were pre-processed using encoding and feature scaling techniques to prepare the data for machine learning analysis. Experimental results demonstrate that the Poly-BoostNet model significantly outperforms existing algorithms including PA, SVM, and ET methods. In classification tasks, the proposed model achieved 100% accuracy, precision, recall, and F1-score, successfully predicting advertisement engagement levels across all classes. Similarly, in regression analysis for customer satisfaction prediction, the model achieved very low error metrics (MAE = 0.0325, RMSE = 0.0455) and an R^2 score of 0.9998, indicating highly accurate predictions. These results confirm that combining polynomial feature extraction with boosting techniques effectively improves model performance. So, the proposed system provides a powerful data-driven solution for businesses to analyse customer behaviour, enhance marketing strategies, and improve customer retention through intelligent analytics.

REFERENCES

[1] Santthosh Saai Reddy Purmani. (2026). Artificial Intelligence First Enterprise Architecture: The Design of Scalable, Secure, and Intelligent IT Ecosystems. American Journal of AI Cyber Computing Management, 6(1(2)), 1–8. https://doi.org/10.64751/ajaccm.2026.v6.n1(2).pp1-8

[2] Patel, S., & Patyrykin, K. (2025). Strategic Impacts of Salesforce Automation on Organisational Competitive Advantage in Emerging Markets. Journal of Posthumanism, 5(12), 357–372. https://doi.org/10.63332/joph.v5i12.3782

[3] Vasagam, M., Kumar, A., & Garg, A. (2026). Learning Execution Plan Embeddings for Multi-Dimensional Query Resource Prediction. IEEE Access.

[4] Imani, M.; Ghaderpour, Z.; Joudaki, M.; Beikmohammadi, A. The Impact of SMOTE and ADASYN on Random Forest and Advanced Gradient Boosting Techniques in Telecom Customer Churn Prediction. In Proceedings of the 2024 10th International Conference on Web Research (ICWR), Tehran, Iran, 24–25 April 2024.

[5] Kalae, U. K. (2021). Enhancing data analytics and reporting efficiency using Power BI and SQL in cloud computing environments. Journal of Computational Analysis and Applications, 29(6), 2021. https://doi.org/10.48047/jocaaa.2021.29.06.48

[6] Poojari, R. Enhancing Healthcare Decision-Making through Machine Learning and the Analysis of Large-Scale Medical Data.

[7] Reddy, S. K. R. Developing a Modular AI Framework to Enhance Scalability and Personalization in Next-Generation Reward Platforms.

[8] Prodduturi, S. M. K. To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code.

[9] Gaddam, S. From Fixed Specifications to Self-Adapting Systems: A Machine Learning Perspective on Software Engineering.

[10] Explainable AI Framework for Policy-Compliant Anomaly Detection in Data Pipelines. (2025). Inaternational Journal of Communication Networks and Information Security, 16(4). https://doi.org/10.48047/ijcnis.16.4.2111

[11] Lee, N.T.; Lee, H.C.; Hsin, J.; Fang, S.H. Prediction of customer behavior changing via a hybrid approach. IEEE Open J. Comput. Soc. 2023, 5, 27–38.

[12] Wang, X.; Xie, L.; Wang, H.; Xing, X.; Wan, W.; Wu, Z.; Ma, X.; Li, Q. Deciphering Explicit and Implicit Features for Reliable, Interpretable; Actionable User Churn Prediction in Online Video Games. IEEE Trans. Vis. Comput. Graph. 2024, 31, 5990–6007.

[13] Babak, A.; Hosseini, S.H. Unveiling the Power of Social Influence: A Machine Learning Framework for Churn Prediction with Network Analysis. IEEE Access 2024, 12, 71271–71285.

[14] Šimović, P.P.; Chen, C.Y.T.; Sun, E.W. Classifying the variety of customers' online engagement for churn prediction with a mixed-penalty logistic regression. Comput. Econ. 2023, 61, 451–485.

[15] Jakob, R.; Lepper, N.; Fleisch, E.; Kowatsch, T. Predicting early user churn in a public digital weight loss intervention. In Proceedings of the CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024.